

# On the Power of Invariant Tests for Hypotheses on a Covariance Matrix

David Preinerstorfer and Benedikt M. Pötscher\*

First version: February 2013

Second version: March 2014

This version: December 2014

## Abstract

The behavior of the power function of autocorrelation tests such as the Durbin-Watson test in time series regressions or the Cliff-Ord test in spatial regression models has been intensively studied in the literature. When the correlation becomes strong, Krämer (1985) (for the Durbin-Watson test) and Krämer (2005) (for the Cliff-Ord test) have shown that the power can be very low, in fact can converge to zero, under certain circumstances. Motivated by these results, Martellosio (2010) set out to build a general theory that would explain these findings. Unfortunately, Martellosio (2010) does not achieve this goal, as a substantial portion of his results and proofs suffer from serious flaws. The present paper now builds a theory as envisioned in Martellosio (2010) in a fairly general framework, covering general invariant tests of a hypothesis on the disturbance covariance matrix in a linear regression model. The general results are then specialized to testing for spatial correlation and to autocorrelation testing in time series regression models. We also characterize the situation where the null and the alternative hypothesis are indistinguishable by invariant tests.

AMS Mathematics Subject Classification 2010: 62F03, 62G10, 62H11, 62H15, 62J05.

Keywords: power function, invariant test, autocorrelation, spatial correlation, zero-power trap, indistinguishability, Durbin-Watson test, Cliff-Ord test.

## 1 Introduction

Testing hypotheses on the covariance matrix of the disturbances in a regression model is an important problem in econometrics and statistics, a prime example being testing the hypothesis of uncorrelatedness of the disturbances. Two particularly important cases are (i) testing for autocorrelation in time series regressions and (ii) testing for spatial autocorrelation in spatial models; for an overview see King (1987) and Anselin (2001). For testing autocorrelation in time series regressions the most popular test is probably the Durbin-Watson test. While low power of this test against highly correlated alternatives in some instances had been noted earlier by Tillman (1975) and King (1985), Krämer (1985) seems to have been the first to show that the limiting power of

---

\*We thank Federico Martellosio and two referees for helpful comments on an earlier version of the paper. Address for correspondence: Department of Statistics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria. E-mail: {david.preinerstorfer, benedikt.poetscher}@univie.ac.at

the Durbin-Watson test as autocorrelation goes to one can actually be zero. This phenomenon has become known as the *zero-power trap*. The work by Krämer (1985) has been followed up and extended in the context of testing against autoregressive disturbances of order one in Zeisel (1989), Krämer and Zeisel (1990), and Löbus and Ritter (2000); see also Small (1993) and Bartels (1992). Loosely speaking, these results show that the power of the Durbin-Watson test (and of a class of related tests) typically converges to either one or zero (depending on whether a certain observable quantity is below or above a threshold) as the strength of autocorrelation increases, provided that there is no intercept in the regression (in the sense that the vector of ones is not in the span of the regressor matrix); in case an intercept is in the regression, the limit is typically neither zero nor one. Some of these results were extended in Kleiber and Krämer (2005) to the case where the Durbin-Watson test is used, but the disturbances are fractionally integrated. In the context of spatial regression models Krämer (2005) showed that the Cliff-Ord test can similarly be affected by the zero-power trap. Martellosio (2010) set out to build a general theory for power properties of tests of a hypothesis on the covariance matrix of the disturbances in a linear regression, that would also uncover the mechanism responsible for the phenomena observed in the before-cited literature. While the intuition behind the general results in Martellosio (2010) is often correct, the results themselves and/or their proofs are not. For example, the main result (Theorem 1 in Martellosio (2010)), on which much of that paper rests, has some serious flaws: Parts of the theorem are incorrect, and the proofs of the correct parts are substantially in error. In particular, the proof in Martellosio (2010) is based on a "concentration" effect, which, however, is simply not present in the setting of the proof of Theorem 1 in Martellosio (2010), as the relevant distributions "stretch out" rather than "concentrate". This has already been observed in Mynbaev (2012), where a way to circumvent the problems was suggested. Mynbaev's approach, which is based on the "stretch-out effect", is somewhat cumbersome in that it requires the development of tools dealing with the "stretch-out effect"; furthermore, the treatment in Mynbaev (2012) is given only for a subclass of the tests considered in Martellosio (2010) and under more restrictive distributional assumptions than in Martellosio (2010).

In the present paper we now build a theory as envisioned in Martellosio (2010) at an even more general level. In particular, we allow for general invariant tests including randomized ones, we employ weaker conditions on the underlying covariance model as well as on the distributions of the disturbances (e.g., we even allow for distributions that are not absolutely continuous). One aspect of our theory is to show how invariance of the tests considered can be used to convert Martellosio's intuition about the "concentration" effect into a precise mathematical argument. Furthermore, advantages of this approach over the approach in Mynbaev (2012) are that (i) standard weak convergence arguments can be used (avoiding the need for new tools to handle the "stretch-out" effect), (ii) more general classes of tests can be treated, and (iii) much weaker distributional assumptions are required. The general theory built in this paper is then applied to tests for spatial autocorrelation, which, in particular, leads to correct versions of the results in Martellosio (2010) that pertain to spatial models.<sup>1</sup> A further contribution of the present paper is a characterization of the situation where no invariant test can distinguish the null hypothesis of no correlation from the alternative. This characterization helps to explain, and provides a unifying framework for, phenomena observed in Kadiyala (1970), Arnold (1979), Kariya (1980), Martellosio (2010), and Martellosio (2011b).

The paper is organized as follows: After laying out the framework in Section 2.1, the general theory is developed in Section 2.2. The main results are Theorems 2.7, 2.16, and 2.18. Theorem

---

<sup>1</sup>This involves more than just providing a correct version of Theorem 1, the main result in Martellosio (2010), and is not undertaken in Mynbaev (2012), see his Remark 2.12.

2.7, specialized to nonrandomized tests, shows that under appropriate assumptions the power of an invariant test converges to 0 or 1 as the "boundary" of the alternative is approached. The limit is 0 or 1 depending on whether a certain observable vector  $e$  (the "concentration direction" of the underlying covariance model) belongs to the complement of the closure or to the interior of the rejection region of the test. This result constitutes a generalization of the correct parts of Theorem 1 in Martellosio (2010) (the proofs of which in Martellosio (2010) are incorrect). Theorems 2.16 and 2.18 deal with the case where the concentration direction  $e$  belongs to the boundary of the rejection region, a case excluded from Theorem 2.7, thus providing correct versions of the incorrect part of Theorem 1 in Martellosio (2010). The general results obtained in Theorems 2.7, 2.16, and 2.18 are then specialized in Section 2.2.3 to the important class of tests based on test statistics that are ratios of quadratic forms. The relationship between test size and the zero-power trap is discussed in Section 2.2.4, before indistinguishability of the null and alternative hypothesis by invariant tests is characterized in Section 2.3. Extensions of the general theory are discussed in Section 3; in particular, we discuss ways of relaxing the distributional assumptions. Section 4 is devoted to applying the general theory to testing for spatial correlation, while Section 5 contains an application to testing for autocorrelation in time series regression models. Whereas the problems with Theorem 1 in Martellosio (2010) are discussed in Section 2.2 as well as in Appendix A, problems with a number of other results in Martellosio (2010) are dealt with in Appendix B. Proofs can be found in Appendices C and D. Some auxiliary results are collected in Appendix E.

## 2 The behavior of the power function: general theory

### 2.1 Framework

As in Martellosio (2010), we consider the problem of testing a hypothesis on the covariance matrix of the disturbance vector in a linear regression model. Given parameters  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and  $\rho \in [0, a)$ , where  $a$  is some prespecified positive real number, the model is

$$\mathbf{y} = X\beta + \mathbf{u}, \quad (1)$$

where  $X \in \mathbb{R}^{n \times k}$  is a non-stochastic matrix of rank  $k$  with  $0 \leq k < n$  and  $n \geq 2$ . [In case  $k = 0$  we identify  $\mathbb{R}^{n \times k}$ , the space of real  $n \times k$  matrices, with  $\{0\} \subseteq \mathbb{R}^n$  and  $\mathbb{R}^k$  with  $\{0\} \subseteq \mathbb{R}$ .] The disturbance vector  $\mathbf{u}$  is assumed to be an  $n \times 1$  random vector with mean zero and covariance matrix  $\sigma^2 \Sigma(\rho)$ , where  $\Sigma(\cdot)$  is a *known* function from  $[0, a)$  to the set of symmetric and positive definite  $n \times n$  matrices. Without loss of generality (w.l.o.g.)  $\Sigma(0)$  is assumed to be the identity matrix  $I_n$ . [The case  $a = \infty$  can be immediately reduced to the case  $a < \infty$  considered here by use of a transformation like  $\arctan(\rho)$ .] We assume furthermore that, given  $\beta$ ,  $\sigma$ , and  $\rho$ , the distribution of  $\mathbf{u}$  is completely specified (but see Remark 3.2 in Section 3 for a relaxation of this assumption). Note that this does not imply in general that the distribution of  $\sigma^{-1} \Sigma^{-1/2}(\rho) \mathbf{u}$  is independent of  $\rho$ ,  $\sigma$ , and  $\beta$  (although this will often be the case in important examples). In contrast to Martellosio (2010) we do not impose any further assumptions on the distribution of  $\mathbf{u}$  at this stage (see Remark 2.1 below for a discussion of the additional assumptions in Martellosio (2010)). All additional distributional assumptions needed later will be stated explicitly in the theorems.

Under the preceding assumptions, model (1) induces a *parametric* family of distributions

$$\mathfrak{P} = \{P_{\beta, \sigma, \rho} : \beta \in \mathbb{R}^k, 0 < \sigma < \infty, \rho \in [0, a)\} \quad (2)$$

on the sample space  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  where  $P_{\beta, \sigma, \rho}$  stands for the distribution of  $\mathbf{y}$  under the given parameters  $\beta$ ,  $\sigma$ , and  $\rho$ , and where  $\mathcal{B}(\mathbb{R}^n)$  denotes the Borel  $\sigma$ -field on  $\mathbb{R}^n$ . The expectation operator with respect to (w.r.t.)  $P_{\beta, \sigma, \rho} \in \mathfrak{P}$  shall be denoted by  $E_{\beta, \sigma, \rho}$ . If  $M$  is a Borel-measurable mapping from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ , we shall denote by  $P_{\beta, \sigma, \rho} \circ M$  the pushforward measure of  $P_{\beta, \sigma, \rho}$  under  $M$ , which is defined on  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ . As usual, a Borel-set  $A$  will be said to be a  $\mathfrak{P}$ -null set if it is a null set relative to every element of  $\mathfrak{P}$ .

**Remark 2.1.** (*Comments on assumptions in Martellosio (2010)*) (i) In Martellosio (2010), p.154, additional assumptions on the distribution of  $\mathbf{u}$  are imposed: for example, it is assumed that  $\mathbf{u}$  possesses a density which is positive everywhere on  $\mathbb{R}^n$ , is larger at 0 than anywhere else, and satisfies a continuity property (the meaning of which is not completely transparent). These assumptions are in general stronger than what is needed; for example, as we shall see, some of our results even hold for discretely distributed errors.

(ii) In Martellosio (2010) it is furthermore implicitly assumed that for fixed  $\rho$ , the distribution of  $\sigma^{-1}\mathbf{u}$  (or, equivalently, the distribution of  $\sigma^{-1}\Sigma^{-1/2}(\rho)\mathbf{u}$ ) does not depend on  $\beta$  and  $\sigma$ . This becomes apparent on p. 156, where it is claimed that the testing problem under consideration is invariant w.r.t. the group  $G_X$  (defined below) in the sense of Lehmann and Romano (2005). In fact, Martellosio (2010) appears to even assume implicitly that the distribution of  $\sigma^{-1}\Sigma^{-1/2}(\rho)\mathbf{u}$  is independent of *all* the parameters  $\beta$ ,  $\sigma$ , and  $\rho$ ; cf., e.g., the first line in the proof of Theorem 1 on p. 182 in Martellosio (2010).

We consider the problem of testing  $\rho = 0$  against  $\rho > 0$ . More precisely, the null hypothesis and the alternative hypothesis are given by

$$H_0 : \rho = 0, \beta \in \mathbb{R}^k, 0 < \sigma < \infty \text{ against } H_1 : \rho > 0, \beta \in \mathbb{R}^k, 0 < \sigma < \infty, \quad (3)$$

with the implicit understanding that always  $\rho \in [0, a)$ . We note that typically one would impose an additional (identifiability) condition such as, e.g.,  $\sigma^2\Sigma(\rho) \neq \tau^2\Sigma(0)$  for every  $\rho > 0$  and every  $0 < \sigma, \tau < \infty$  in order to ensure that  $H_0$  and  $H_1$  are disjoint, and hence that the test problem is meaningful.<sup>2</sup> The results on the power behavior as  $\rho \rightarrow a$  in the present paper are valid without any such explicit identifiability condition, but note that one of the basic assumptions (Assumption 1) underlying most of the results automatically implies that  $\sigma^2\Sigma(\rho) \neq \tau^2\Sigma(0)$  for every  $0 < \sigma, \tau < \infty$  holds at least for  $\rho > 0$  in a neighborhood of  $a$ .

A (randomized) test is a Borel-measurable function  $\varphi$  from the sample space  $\mathbb{R}^n$  to  $[0, 1]$ , and a non-randomized test is the indicator function of a set  $\Phi \in \mathcal{B}(\mathbb{R}^n)$ , the rejection region. A test statistic is a Borel-measurable function  $T : \mathbb{R}^n \rightarrow \mathbb{R}$  which, together with a critical value  $\kappa \in \mathbb{R}$ , gives rise to a rejection region  $\{y \in \mathbb{R}^n : T(y) > \kappa\}$ .<sup>3</sup> Note that the tests (rejection regions, test statistics, critical values) may depend on the sample size  $n$  as well as on the design matrix  $X$ , but typically we shall not show this in the notation. Recall that the size of a test  $\varphi$  is given by  $\sup_{\beta \in \mathbb{R}^k} \sup_{0 < \sigma < \infty} E_{\beta, \sigma, 0}(\varphi)$ , i.e., is the supremal rejection probability under the null.

We shall also use the following terminology and notation: Random vectors and matrices will always be denoted by boldface letters. All matrices considered will be real matrices. The transpose of a matrix  $A$  is denoted by  $A'$ . The space spanned by the columns of  $A$  is denoted by  $\text{span}(A)$ .

<sup>2</sup>Of course, even if  $\sigma^2\Sigma(\rho) = \tau^2\Sigma(0)$  holds for some  $\sigma > 0$ ,  $\tau > 0$  and some  $\rho > 0$ , there may still be additional identifying information present in the distributions that goes beyond the information contained in first and second moments.

<sup>3</sup>The case of a test statistic  $S$  taking values in the extended real line can be easily accommodated in our framework by passing from  $S$  to a real-valued test statistic such as, e.g.,  $T = \arctan(S)$ .

Given a linear subspace  $L$  of  $\mathbb{R}^n$ , the symbol  $\Pi_L$  denotes orthogonal projection onto  $L$ , and  $L^\perp$  denotes the orthogonal complement of  $L$ . Given an  $n \times m$  matrix  $Z$  of rank  $m$  with  $0 \leq m < n$ , we denote by  $C_Z$  a matrix in  $\mathbb{R}^{(n-m) \times n}$  such that  $C_Z C_Z' = I_{n-m}$  and  $C_Z' C_Z = \Pi_{\text{span}(Z)^\perp}$  where  $I_r$  denotes the identity matrix of dimension  $r$ . It is easily seen that every matrix whose rows form an orthonormal basis of  $\text{span}(Z)^\perp$  satisfies these two conditions and vice versa, and hence any two choices for  $C_Z$  are related by premultiplication by an orthogonal matrix. Let  $l$  be a positive integer. If  $A$  is an  $l \times l$  matrix and  $\lambda \in \mathbb{R}$  is an eigenvalue of  $A$  we denote the corresponding eigenspace by  $\text{Eig}(A, \lambda)$ . The eigenvalues of a symmetric matrix  $B \in \mathbb{R}^{l \times l}$  ordered from smallest to largest and counted with their multiplicities are denoted by  $\lambda_1(B), \dots, \lambda_l(B)$ . If  $B$  is a symmetric and nonnegative definite  $l \times l$  matrix, every  $l \times l$  matrix  $A$  that satisfies  $AA' = B$  is called a *square root* of  $B$ ; with  $B^{1/2}$  we denote its unique symmetric and nonnegative definite square root. Note that every square root of  $B$  is of the form  $B^{1/2}U$  for some orthogonal matrix  $U$ . A vector  $x \in \mathbb{R}^l$  is said to be normalized if  $\|x\| = 1$ , where  $\|\cdot\|$  denotes Euclidean norm on  $\mathbb{R}^l$ . The operators  $\text{bd}$ ,  $\text{int}$ , and  $\text{cl}$  shall denote the boundary, interior, and closure of a subset of  $\mathbb{R}^l$  w.r.t. the Euclidean topology. For  $x \in \mathbb{R}^l$  the symbol  $\delta_x$  denotes point mass at  $x$ . Lebesgue measure on  $(\mathbb{R}^l, \mathcal{B}(\mathbb{R}^l))$  shall be denoted by  $\mu_{\mathbb{R}^l}$ , while Lebesgue measure on the Borel subsets of  $(0, \infty)$  is denoted by  $\mu_{(0, \infty)}$ . The uniform probability measure on the Borel subsets of  $S^{n-1}$ , the unit sphere in  $\mathbb{R}^n$ , is denoted by  $\nu_{S^{n-1}}$ . We use  $\text{Pr}$  as a generic symbol for a probability measure, with  $E$  denoting the corresponding expectation operator.

### 2.1.1 Groups of transformations, invariance, and maximal invariants

Suppose that  $G$  is a group of bijective Borel-measurable transformations  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , the group operation being composition. A function  $F$  defined on  $\mathbb{R}^n$  is said to be invariant w.r.t.  $G$  if for every  $y \in \mathbb{R}^n$  and every  $g \in G$  we have  $F(y) = F(g(y))$ . A subset  $A$  of  $\mathbb{R}^n$  is said to be invariant w.r.t.  $G$  if for every  $g \in G$  we have that  $g(A) \subseteq A$ .<sup>4</sup> Of course, invariance of  $F$  implies invariance of  $\{y \in \mathbb{R}^n : F(y) > \kappa\}$ .

Given a matrix  $Z \in \mathbb{R}^{n \times m}$  such that  $0 \leq m < n$  with column rank  $m$ , we will mainly work with the group

$$G_Z = \{g_{\gamma, \theta} : \gamma \in \mathbb{R} \setminus \{0\}, \theta \in \mathbb{R}^m\},$$

where  $g_{\gamma, \theta}$  denotes the mapping  $y \mapsto \gamma y + Z\theta$ . The main reason for concentrating on invariance w.r.t. this group is that the majority of tests for the hypothesis (3) considered in the literature have this invariance property (for  $Z = X$ ). Another reason is that this is also the notion of invariance used in Martellosio (2010). Occasionally we shall consider invariance w.r.t. subgroups of  $G_Z$ , see Remark 2.4.

The following is a maximal invariant w.r.t.  $G_Z$

$$\mathcal{I}_Z(y) = \begin{cases} \langle \Pi_{\text{span}(Z)^\perp} y / \|\Pi_{\text{span}(Z)^\perp} y\| \rangle & \text{if } y \notin \text{span}(Z), \\ 0 & \text{else,} \end{cases}$$

where the function  $\langle \cdot \rangle : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined as follows:  $\langle y \rangle$  equals  $y$  multiplied by the sign of the first nonzero coordinate of  $y$  whenever  $y \neq 0$ , and  $\langle y \rangle = 0$  if  $y = 0$  (see Preinerstorfer and Pötscher (2013), Section 5.1, where the group  $G_Z$  is denoted as  $G(\text{span}(Z))$ ). More generally, let  $\zeta$  be any function from the unit sphere in  $\mathbb{R}^n$  into itself that satisfies  $\zeta(y) = \zeta(-y)$  and has the property

---

<sup>4</sup>The group structure implies that this is equivalent to  $g(A) = A$  for every  $g \in G$ , and thus to invariance of the indicator function of  $A$ .

that  $\zeta(y)$  is collinear with  $y$ ; then defining  $\mathcal{I}_{Z,\zeta}(y)$  in the same way as  $\mathcal{I}_Z(y)$ , but with  $\langle \cdot \rangle$  replaced by  $\zeta$ , provides another maximal invariant w.r.t.  $G_Z$ . Obviously, given any normalized vector  $e$ , we can find a  $\zeta$  as above that is additionally Borel-measurable and is continuous in a neighborhood (in the unit sphere) of  $e$ . For such a  $\zeta$  the maximal invariant  $\mathcal{I}_{0,\zeta}$  w.r.t.  $G_0$  is then continuous in a neighborhood of  $e$  (in  $\mathbb{R}^n$ ), and hence in a neighborhood of  $\lambda e$  for any  $\lambda \neq 0$  (in  $\mathbb{R}^n$ ). Moreover, we can even choose  $\zeta$  to be as before and also to satisfy  $\zeta(e) = e$ .<sup>5</sup> In the following we shall write  $\zeta_e$  for any such  $\zeta$ .<sup>6</sup>

**Remark 2.2.** (i) For any test  $\varphi$  invariant w.r.t.  $G_Z$  we have  $\varphi(y) = \varphi(\mathcal{I}_Z(y)) = \varphi(\mathcal{I}_{Z,\zeta}(y))$  for every  $y \in \mathbb{R}^n$  and  $\zeta$  as above. This is trivial for  $y \in \text{span}(Z)$  since  $\varphi(y) = \varphi(0)$  must hold by invariance. For  $y \notin \text{span}(Z)$  observe that due to invariance we have

$$\varphi(y) = \varphi(\Pi_{\text{span}(Z)^\perp} y) = \varphi(\Pi_{\text{span}(Z)^\perp} y / \|\Pi_{\text{span}(Z)^\perp} y\|) = \varphi(\mathcal{I}_Z(y)) = \varphi(\mathcal{I}_{Z,\zeta}(y)),$$

noting that  $\mathcal{I}_Z(y)$  as well as  $\mathcal{I}_{Z,\zeta}(y)$  are proportional to  $\Pi_{\text{span}(Z)^\perp} y / \|\Pi_{\text{span}(Z)^\perp} y\|$  with a proportionality factor equal to  $\pm 1$ .

(ii) For later use we note the following: if  $\varphi$  is invariant w.r.t.  $G_Z$ , it is also invariant w.r.t.  $G_0$ . Consequently, we have  $\varphi(y) = \varphi(\mathcal{I}_0(y)) = \varphi(\mathcal{I}_{0,\zeta}(y))$  for every  $y \in \mathbb{R}^n$  and  $\zeta$  as above.

**Remark 2.3.** If one assumes that the distribution of  $\sigma^{-1}\mathbf{u}$  does not depend on  $\beta$  and  $\sigma$  (as is, e.g., done in Martellosio (2010), cf. Remark 2.1(ii) above), the power function of any  $G_X$ -invariant test  $\varphi$  is then independent of  $\beta$  and  $\sigma$ ; that is, for every  $\rho \in [0, a)$  we have

$$E_{\beta,\sigma,\rho}(\varphi) = E_{0,1,\rho}(\varphi) \text{ for every } \beta \in \mathbb{R}^k, 0 < \sigma < \infty.$$

If, additionally, all the parameters of the model are identifiable, the test problem (3) is then in fact a  $G_X$ -invariant test problem in the sense of Lehmann and Romano (2005), Chapter 6.

**Remark 2.4.** In Sections 2.3 and 4.3 as well as in Remark 3.3 we shall also consider invariance w.r.t. the subgroups  $G_Z^+ = \{g_{\gamma,\theta} : \gamma > 0, \theta \in \mathbb{R}^m\}$  and  $G_Z^1 = \{g_{1,\theta} : \theta \in \mathbb{R}^m\}$  with associated maximal invariants

$$\mathcal{I}_Z^+(y) = \begin{cases} \Pi_{\text{span}(Z)^\perp} y / \|\Pi_{\text{span}(Z)^\perp} y\| & \text{if } y \notin \text{span}(Z), \\ 0 & \text{else,} \end{cases}$$

and  $\mathcal{I}_Z^1(y) = \Pi_{\text{span}(Z)^\perp} y$ , respectively.

## 2.2 Main results

We now set out to study the behavior of the power function of invariant tests for the testing problem (3) when the parameter  $\rho$  is ‘far away’ from 0, the value of  $\rho$  under the null hypothesis, i.e., when  $\rho$  is close to its upper limit  $a$ . In particular, we are interested in the *limiting power* of such tests  $\varphi$  as  $\rho \rightarrow a$ , i.e., in  $\lim_{\rho \rightarrow a} E_{\beta,\sigma,\rho}(\varphi)$ . For these limits as well as for all other limits where  $\rho \rightarrow a$  it is always implicitly understood that  $\rho \in [0, a)$ , i.e., that one is considering left-hand side limits. [To avoid confusion, we stress that throughout we consider a finite-sample situation, i.e., sample size  $n$

<sup>5</sup>In fact,  $\zeta$  then coincides with the identity in a neighborhood (in the unit sphere) of  $e$ .

<sup>6</sup>On p. 156 of Martellosio (2010) it is claimed that the quantity  $\nu$  defined there is a maximal invariant for the group  $G_X$  (denoted by  $F_X$  in Martellosio (2010)). First note that the author does not spell out how  $\nu$  is defined for  $y \in \text{span}(Z)$  and how  $\text{sgn}(0)$  is to be interpreted. Second, regardless of how one defines  $\nu$  on  $\text{span}(Z)$  and whether one interpretes  $\text{sgn}(0)$  as 0, 1, or  $-1$ , the quantity  $\nu$  is not invariant in general as can be seen from simple examples.

is fixed, and hence the notion of limiting power just introduced has nothing to do with asymptotic power properties where sample size increases to infinity.] To motivate our interest in this problem we consider the following two examples.

**Example 2.1.** (*Testing for positive autocorrelation*) Assume that the disturbances in the regression model (1) follow a Gaussian stationary autoregressive process of order one with autoregressive parameter  $\rho$ . Then the  $(i, j)$ -th element of  $\Sigma(\rho)$  is given by  $(1 - \rho^2)^{-1} \rho^{|i-j|}$  for  $\rho \in [0, 1)$ , i.e.,  $a = 1$ . Unguided intuition may suggest that the power of standard tests like the Durbin-Watson test for testing  $\rho = 0$  versus  $\rho > 0$  is large if  $\rho$  is sufficiently different from zero, and, in particular, if  $\rho$  is close to  $a = 1$ . In fact, this intuition may even suggest that the power of the Durbin-Watson test should approach 1 as  $\rho \rightarrow a = 1$ . However, as already mentioned in the introduction, this intuition is wrong: The limiting power of the Durbin-Watson test can be zero (or one, or a number in  $(0, 1)$ ) depending on the design matrix and the significance level employed (see Krämer (1985), Zeisel (1989), Krämer and Zeisel (1990), and Löbus and Ritter (2000)).  $\square$

**Example 2.2.** (*Testing for spatial autocorrelation*) Assume now that the disturbances in the regression model (1) are Gaussian spatial autoregressive errors of order one. Then under typical assumptions on the spatial weights matrix  $W$  we have

$$\Sigma(\rho) = (I_n - \rho W)^{-1} (I_n - \rho W')^{-1}$$

for  $\rho \in [0, \lambda_{\max}^{-1})$ , i.e.,  $a = \lambda_{\max}^{-1}$ . Here  $\lambda_{\max} > 0$  is a dominant eigenvalue of  $W$ . As in the preceding example, unguided intuition may suggest that the limiting power of standard tests like the Cliff-Ord test for  $\rho \rightarrow a = \lambda_{\max}^{-1}$  is large (e.g., is equal to 1). However, this intuition is again incorrect and the limiting power of the Cliff-Ord test can be zero (or one, or a number in  $(0, 1)$ ) depending on the design matrix, the weights matrix, and the significance level employed (see Krämer (2005)).  $\square$

Our goal is now to develop a coherent theory for deriving the limiting power of invariant tests for the testing problem (3), which allows for more general correlation structures than the ones figuring in the preceding examples and which allows for non-Gaussian distributions. As mentioned in the introduction, an attempt at such a theory has been made in Martellosio (2010) and it is thus appropriate as a starting point to revisit and discuss the main result in that paper: A large part of Martellosio (2010) is devoted to determining the limiting power of non-randomized tests  $\mathbf{1}_\Phi$  as  $\rho \rightarrow a$ , i.e.,  $\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\Phi)$ . Not surprisingly, the limiting behavior of the power function crucially depends on the behavior of the function  $\Sigma$  close to  $a$ . Martellosio (2010) concentrates on situations where  $\Sigma^{-1}(a-) := \lim_{\rho \rightarrow a} \Sigma^{-1}(\rho)$  exists in  $\mathbb{R}^{n \times n}$ , and, in particular, on the case where the rank of  $\Sigma^{-1}(a-)$  equals  $n - 1$ .<sup>7,8</sup> It should be observed that this condition on the function  $\Sigma$  is satisfied in the two examples discussed above. In the following we quote Theorem 1, the main result of Martellosio (2010), which is set in the framework described in Section 2.1 augmented by the additional distributional assumptions of Martellosio (2010), discussed above in Remark 2.1:

<sup>7</sup>The case where  $\Sigma^{-1}(a-)$  exists and is positive definite (equivalently, where the left-hand side limit  $\Sigma(a-)$  of  $\Sigma(\cdot)$  exists and is positive definite) is not the focus of Martellosio (2010) since then the model is typically also well-defined for  $\rho = a$  and the limiting power for  $\rho \rightarrow a$  typically coincides with the power for  $\rho = a$ .

<sup>8</sup>In Martellosio (2010), p. 159, it is claimed that the following three cases are exhaustive: (i)  $\lim_{\rho \rightarrow a} \Sigma(\rho)$  exists and is positive definite; (ii)  $\lim_{\rho \rightarrow a} \Sigma(\rho)$  exists and is singular and (iii)  $\lim_{\rho \rightarrow a} \Sigma^{-1}(\rho)$  exists and is singular. This does not provide an exhaustive description of possible cases, as there exist functions  $\Sigma$  such that neither  $\lim_{\rho \rightarrow a} \Sigma(\rho)$  nor  $\lim_{\rho \rightarrow a} \Sigma^{-1}(\rho)$  exist. Let  $n = 2$  and define  $\Sigma(\rho)$  as a diagonal matrix with diagonal  $(1 - \rho, (1 - \rho)^{-1})$  for  $\rho \in [0, 1)$ . Clearly, both  $\Sigma(\rho)$  and its inverse do not converge as  $\rho \rightarrow 1$ .

”Consider an invariant critical region  $\Phi$  for testing  $\rho = 0$  against  $\rho > 0$  in model (1). Assume that  $\Sigma(\rho)$  is positive definite as  $\rho \rightarrow a$ <sup>9</sup>, and that  $\text{rank}(\Sigma^{-1}(a)) = n - 1$ . The limiting power of  $\Phi$  as  $\rho \rightarrow a$  is:

- 1 if  $f_1(\Sigma^{-1}(a)) \in \text{int}(\Phi)$ ;
- in  $(0, 1)$  if  $f_1(\Sigma^{-1}(a)) \in \text{bd}(\Phi)$ ; or
- 0 if  $f_1(\Sigma^{-1}(a)) \notin \text{cl}(\Phi)$ .”

From now on we shall refer to this theorem of Martellosio (2010) as MT1. A few comments are in order: First, the notion of invariance used in the quote is invariance w.r.t.  $G_X$ . Second, observe that even if  $\Sigma(\rho)$  is well-defined for  $\rho = a$  (which need not be the case in general), the statement  $\text{rank}(\Sigma^{-1}(a)) = n - 1$  as given in the formulation of MT1 can obviously never be satisfied. To give meaning to the above quote, the symbol  $\Sigma^{-1}(a)$  needs to be interpreted as  $\Sigma^{-1}(a-)$  throughout; this also becomes transparent from the proof in Martellosio (2010). Third, the symbol  $f_1(A)$  in the above quote denotes a normalized eigenvector of a symmetric matrix  $A$  pertinent to  $\lambda_1(A)$ , the smallest eigenvalue of  $A$ . Note that  $\lambda_1(\Sigma^{-1}(a-)) = 0$  due to the rank assumption in the quote. Furthermore, by the rank assumption  $f_1(\Sigma^{-1}(a-))$  is uniquely determined up to a sign change; because  $\Phi$  is  $G_X$ -invariant, the validity of conditions like  $f_1(\Sigma^{-1}(a-)) \in \text{int}(\Phi)$  therefore does not depend on the choice of sign. Fourth, if  $\Phi$  or its complement is a (non-empty)  $\mu_{\mathbb{R}^n}$ -null set, then the second claim of MT1 can obviously not hold. While these cases are unfortunately not ruled out explicitly in the statement of MT1 (which may lead to confusion among some readers), it should be noted that such cases are implicitly excluded in Martellosio (2010), as the author considers only  $G_X$ -invariant rejection regions  $\Phi$  that have size strictly between zero and one, cf. Martellosio (2010), p. 157. [Note that under the distributional assumptions in Martellosio (2010), cf. Remark 2.1 above,  $G_X$ -invariance of  $\Phi$  implies that the size of  $\Phi$  is given by  $P_{0,1,0}(\Phi)$  and that this is 0 (or 1) precisely if  $\Phi$  (or its complement) is a  $\mu_{\mathbb{R}^n}$ -null set.]

Even with the just discussed appropriate interpretations, the second claim in MT1 is incorrect (cf. also Mynbaev (2012)), and the proofs of the correct parts (i.e., claims 1 and 3) are in error. Counterexamples to the second claim in MT1 are provided in Examples A.1 and A.2 in Appendix A. A discussion of the mistakes in the proof of the correct parts of MT1 is also given in Appendix A. The following section provides a generalization of the (correct) claims 1 and 3 in MT1, whereas correct versions of the (incorrect) second claim in MT1 are provided in Section 2.2.2.

### 2.2.1 A generalization of the first and third claim in Theorem 1 in Martellosio (2010)

The proof of MT1 given in Martellosio (2010) rests on a ”concentration effect” to occur in the distributions  $P_{\beta,\sigma,\rho}$  as  $\rho \rightarrow a$ , namely that these distributions (for fixed  $\beta$  and  $\sigma$ ) converge (in an appropriate sense) as  $\rho \rightarrow a$  to a distribution concentrated on a one-dimensional subspace. However, as discussed in Appendix A, this concentration effect simply does not occur in the way as claimed in Martellosio (2010) (cf. also Mynbaev (2012)). In fact, the direct opposite happens: the distributions  $P_{\beta,\sigma,\rho}$  stretch out, i.e., all of the mass ”escapes to infinity”. As we shall now show, the problem can, however, be fixed: The crucial observation is that, while rescaling the data has no effect on the rejection probability of  $G_X$ -invariant tests, an appropriate rescaling can enforce

---

<sup>9</sup>Positive definiteness is always assumed in Martellosio (2010) for  $\rho \in [0, a)$ , hence this assumption seems to be superfluous.



the desired concentration effect. Formalizing this observation will lead us to Theorem 2.7, which provides a generalization of the first and third claim of MT1 under even weaker distributional assumptions than the ones used in MT1; in addition, this theorem will also cover randomized tests. For a discussion and some intuition regarding the concentration effect in a different setting see Preinerstorfer and Pötscher (2013), Section 5.2.<sup>10</sup> We shall make use of the following assumption on the function  $\Sigma$  which is weaker than the rank assumption in MT1.

**Assumption 1.**  $\lambda_n^{-1}(\Sigma(\rho))\Sigma(\rho) \rightarrow ee'$  as  $\rho \rightarrow a$  for some  $e \in \mathbb{R}^n$ .

Note that the vector  $e$  in Assumption 1 is necessarily normalized and will be called *concentration direction* of the underlying covariance model. That this assumption is indeed weaker than the assumption of a one-dimensional kernel of  $\Sigma^{-1}(a-)$  made in MT1 is shown in the following lemma.<sup>11</sup> Recall that when writing  $\Sigma^{-1}(a-)$  we always implicitly assume that this limit exists in  $\mathbb{R}^{n \times n}$ .

**Lemma 2.5.** *If the normalized vector  $e$  spans the kernel of  $\Sigma^{-1}(a-)$ , then  $\lambda_n^{-1}(\Sigma(\rho))\Sigma(\rho) \rightarrow ee'$  as  $\rho \rightarrow a$ .*

The converse is not true as shown in the subsequent example. This shows that Assumption 1 underlying Theorem 2.7 given below is strictly weaker than the assumption of a one-dimensional kernel of  $\Sigma^{-1}(a-)$  underlying MT1.

**Example 2.3.** For  $\rho \in [0, 1)$  let  $\Sigma(\rho)$  be a  $2 \times 2$  diagonal matrix with diagonal entries 1 and  $1 - \rho$ . Then the largest eigenvalue of  $\Sigma(\rho)$  equals one and  $\lambda_n^{-1}(\Sigma(\rho))\Sigma(\rho)$  converges to  $ee'$  as  $\rho \rightarrow 1$ , where  $e = (1, 0)'$ . But the limit of  $\Sigma^{-1}(\rho)$  for  $\rho \rightarrow 1$  does obviously not exist. Another example, where  $\lambda_n^{-1}(\Sigma(\rho))\Sigma(\rho) \rightarrow ee'$  and for which the limit of  $\Sigma^{-1}(\rho)$  for  $\rho \rightarrow 1$  exists, but does not have a one-dimensional kernel, is provided by the  $2 \times 2$  diagonal matrix with diagonal entries  $(1 - \rho)^{-1}$  and  $(1 - \rho)^{-1/2}$ . In this case the limit  $\Sigma^{-1}(1-)$  exists and equals the zero matrix.  $\square$

For  $\xi \in \mathbb{R}^n$  and  $\delta \in \mathbb{R} \setminus \{0\}$  let  $M_{\xi, \delta}$  denote the mapping  $y \mapsto \delta^{-1}(y - \xi)$  from  $\mathbb{R}^n$  to  $\mathbb{R}^n$ . We now introduce the following high-level assumption on  $\mathfrak{P}$  which will be seen to be satisfied under the assumptions in Martellosio (2010) underlying MT1. Simple sufficient conditions for this assumption that are frequently satisfied are discussed below.

**Assumption 2.** For every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and every sequence  $\rho_m \in [0, a)$  converging to  $a$ , every weak accumulation point  $P$  of

$$P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma} \quad (4)$$

satisfies  $P(\{0\}) = 0$ .

The measure in (4) will in general *not* coincide with  $P_{0, \lambda_n^{-1/2}(\Sigma(\rho_m)), \rho_m}$ . However, in the important special case, where the distribution of  $\sigma^{-1}\mathbf{u}$  does not depend on  $\beta$  and  $\sigma$  (cf. Remark 2.3), these two measures will indeed coincide. We furthermore note that in view of Lemma C.1 in Appendix C the sequence in (4) is automatically tight whenever Assumption 1 is satisfied.

Before we present our generalizations of the first and third claim in MT1 we provide simple sufficient conditions for the high-level Assumption 2. To this end we introduce the following assumption on  $\mathfrak{P}$  that is clearly satisfied in many examples.

<sup>10</sup>In the setting of Preinerstorfer and Pötscher (2013) no rescaling is needed to achieve the concentration effect.

<sup>11</sup>The proof idea is also used in the proof of Lemma E.4 in Martellosio (2010) in the special case of a SAR(1) model. See also Lemma 3.3 in Martellosio (2011a) and its proof.

**Assumption 3.** There exists an  $n \times 1$  random vector  $\mathbf{z}$  with mean zero and covariance matrix  $I_n$  such that for every  $\beta \in \mathbb{R}^k$ , every  $0 < \sigma < \infty$ , and every  $\rho \in [0, a)$  the distribution  $P_{\beta, \sigma, \rho}$  is induced by model (1) with  $\mathbf{u}$  having the same distribution as  $\sigma L(\rho)\mathbf{z}$  and where the matrices  $L(\rho)$  satisfy  $L(\rho)L'(\rho) = \Sigma(\rho)$ .<sup>12</sup>

Important examples of families  $\mathfrak{P}$  satisfying Assumption 3 are provided by elliptically symmetric families. Here  $\mathfrak{P}$  is said to be an *elliptically symmetric family* if it satisfies Assumption 3 and  $\mathbf{z}$  is spherically symmetric, i.e., the distributions of  $U\mathbf{z}$  and  $\mathbf{z}$  are the same for every orthogonal matrix  $U$ .<sup>13</sup> Obviously, if  $\mathfrak{P}$  is an elliptically symmetric family, we can assume without loss of generality that  $L(\rho) = \Sigma^{1/2}(\rho)$  in Assumption 3 (because any  $L(\rho)$  satisfies  $L(\rho) = \Sigma^{1/2}(\rho)U(\rho)$  for some orthogonal matrix  $U(\rho)$ ). Furthermore, recall from Remark 2.1 that Martellosio (2010) implicitly imposes Assumption 3 (with  $L(\rho) = \Sigma^{1/2}(\rho)$ ) and more. Sufficient conditions for Assumption 2 are now as follows.

**Proposition 2.6.** <sup>14</sup> Suppose Assumptions 1 and 3 are satisfied.

1. If  $L(\rho) = \Sigma^{1/2}(\rho)$  and  $\Pr(e'\mathbf{z} = 0) = 0$  hold for  $e$  as in Assumption 1 and for  $L(\cdot)$  and  $\mathbf{z}$  as in Assumption 3, then  $\mathfrak{P}$  satisfies Assumption 2.
2. If the distribution of  $\mathbf{z}$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ , then  $\mathfrak{P}$  satisfies Assumption 2. More generally, if  $\Pr(\mathbf{z} = 0) = 0$  and the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  is absolutely continuous w.r.t. the uniform distribution  $v_{S^{n-1}}$  on the unit sphere  $S^{n-1}$ , then  $\mathfrak{P}$  satisfies Assumption 2.

We note that Part 1 of the preceding proposition shows that Assumption 2 also allows for families of discrete distributions. In some contexts (e.g., spatial regression models) it is convenient to avoid the assumption  $L(\rho) = \Sigma^{1/2}(\rho)$  made in Part 1. Part 2 shows that this assumption can indeed be avoided at the cost of introducing additional conditions on the distribution of  $\mathbf{z}$ . That the assumptions for the second statement in Part 2 are indeed weaker than the assumptions for the first statement in Part 2 follows from Lemma E.1 in Appendix E.

We are now ready to present and prove a generalization of the first and third claim in MT1. The result is stated for possibly randomized tests.

**Theorem 2.7.** Suppose Assumptions 1 and 2 are satisfied and let  $\varphi$  be a test that is invariant w.r.t.  $G_X$  and is continuous at  $e$ , where  $e$  is as in Assumption 1. Then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have that  $E_{\beta, \sigma, \rho}(\varphi) \rightarrow \varphi(e)$  as  $\rho \rightarrow a$ .

In the next remark we discuss why Theorem 2.7 contains the first and the third claim of MT1 as special cases.

**Remark 2.8.** (i) First observe that in light of Lemma 2.5, Proposition 2.6, and Remark 2.1 the assumptions of Theorem 2.7 are weaker than the assumptions in MT1. Second, under the assumptions of MT1 the vector  $e$  coincides with  $f_1(\Sigma^{-1}(a-))$  in MT1 (possibly up to an irrelevant sign). Third, if  $\varphi$  in Theorem 2.7 is specialized to the indicator function of a rejection region  $\Phi$  that is invariant w.r.t.  $G_X$ , the above theorem reduces to:

<sup>12</sup>Note, in particular, that the distribution of  $\mathbf{z}$  is independent of  $\beta$ ,  $\sigma$ , and  $\rho$ .

<sup>13</sup>The notion of an elliptically symmetric family implies elliptical symmetry of its elements, but is stronger (as the distribution of  $\mathbf{z}$  in Assumption 3 is not allowed to vary with the parameters).

<sup>14</sup>Inspection of the proof shows that, more generally, Assumptions 1 and 3 imply Assumption 2 as soon as  $\Pr(e'U\mathbf{z} = 0) = 0$  holds for any orthogonal matrix that arises as an accumulation point of  $\Sigma^{-1/2}(\rho)L(\rho)$  for  $\rho \rightarrow a$ .

- If  $e \in \text{int}(\Phi)$ , then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have  $\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\Phi) = 1$ , and
- if  $e \notin \text{cl}(\Phi)$ , then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have  $\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\Phi) = 0$ .

To see this simply observe that  $\varphi = \mathbf{1}_\Phi$  is continuous at  $e$  if and only if  $e \notin \text{bd}(\Phi)$ .

(ii) If  $e \in \text{bd}(\Phi)$ , Theorem 2.7 is not applicable as it stands because  $\mathbf{1}_\Phi$  is then not continuous at  $e$ . However, in some cases Theorem 2.7 can be used in an indirect way as follows: suppose the rejection region  $\Phi$  can be modified into an ‘equivalent’ rejection region  $\Phi^*$  (in the sense that  $\Phi$  and  $\Phi^*$  differ only by a  $\mathfrak{P}$ -null set) such that now  $e \notin \text{bd}(\Phi^*)$  holds. As  $\Phi$  and  $\Phi^*$  give rise to the same rejection probabilities, we can therefore obtain the limits of the rejection probabilities of  $\Phi$  by applying Theorem 2.7 to  $\Phi^*$ . More generally, suppose  $\varphi$  is a test that is equal to a test  $\varphi^*$  outside of a  $\mathfrak{P}$ -null set and suppose that  $\varphi^*$  satisfies the assumptions of Theorem 2.7. As  $\varphi$  and  $\varphi^*$  have the same rejection probabilities, we can conclude that  $E_{\beta, \sigma, \rho}(\varphi) \rightarrow \varphi^*(e)$  as  $\rho \rightarrow a$ . [Of course, a simple sufficient condition for  $\mathfrak{P}$ -almost everywhere equality of  $\varphi = \varphi^*$  is that  $\mathfrak{P}$  is dominated by a measure  $\nu$ , say, and  $\varphi = \varphi^*$  holds  $\nu$ -almost everywhere.]

**Remark 2.9.** Theorem 2.7 applies to  $G_X$ -invariant tests. Such tests have a natural justification if the underlying test problem is invariant under  $G_X$  itself (which is not in general required in Theorem 2.7). Recall from Remark 2.3 that the test problem (3) is invariant under  $G_X$  provided the distribution of  $\sigma^{-1}\mathbf{u}$  does not depend on  $\beta$  and  $\sigma$  (which is, e.g., the case under Assumption 3) and the parameters of the model are identified.

### 2.2.2 Correct versions of the second claim in Theorem 1 in Martellosio (2010)

As noted before, the second claim in MT1 is incorrect in general and counterexamples to this claim are provided in Examples A.1 and A.2 in Appendix A. In this section we now aim at establishing correct versions of this result under appropriate assumptions. Theorem 2.16 below will, in particular, provide an explicit expression for the limiting power in case  $e \in \text{span}(X)$ . Since  $\text{span}(X)$  turns out to always be a subset of the boundary of any critical region  $\Phi$  ( $\neq \emptyset, \mathbb{R}^n$ ) that is invariant under  $G_X$  (cf. Proposition 2.11 below), Theorem 2.16 can thus be seen as a partial substitute for the second claim in MT1 (recall that under the assumptions in Martellosio (2010)  $e$  reduces to  $f_1(\Sigma^{-1}(a-))$ ). Furthermore, in the important special case where the critical region is of the form  $\Phi = \{y \in \mathbb{R}^n : T(y) > \kappa\}$ , with  $T$  invariant under  $G_X$  and satisfying some regularity conditions, Theorem 2.18 below will provide explicit expressions for the limiting power in case  $T(e) = \kappa$ . For an important subclass of  $G_X$ -invariant test statistics  $T$  (including certain ratios of quadratic forms in  $y$ ), Proposition 2.11 below will show that  $\text{bd}(\Phi) = \text{span}(X) \cup \{y \in \mathbb{R}^n : T(y) = \kappa\}$  holds (provided  $\emptyset \neq \Phi \neq \mathbb{R}^n$ ). Hence, for this subclass of tests, an application of Theorems 2.16 and 2.18 together provides a substitute for the second claim in MT1 (because then  $e \in \text{bd}(\Phi)$  amounts to  $e \in \text{span}(X)$  or  $T(e) = \kappa$ ).<sup>15</sup> Before we can give these results we need to study the structure of  $\text{bd}(\Phi)$  for  $\Phi$  a  $G_X$ -invariant rejection region.

**On the structure of the boundary of  $G_X$ -invariant rejection regions.** Martellosio (2010), Footnote 9, points out that a  $G_X$ -invariant rejection region  $\Phi$  always satisfies  $\text{span}(X) \subseteq \text{bd}(\Phi)$  provided its size is neither zero nor one. Even if the rejection region is assumed to be of the form  $\Phi = \{y \in \mathbb{R}^n : T(y) > \kappa\}$ , then – contrary to claims in Martellosio (2010) – not much more can be said about the boundary  $\text{bd}(\Phi)$  in general. This is discussed in the subsequent remark. In the

<sup>15</sup>For the discussion in this paragraph we have implicitly assumed that the vector  $e$  in Assumption 1 and Assumption 4 is the same; cf. Remark 2.15 below.

proposition following the remark we show how Martellosio (2010)'s claims, which are incorrect in general, can be saved if additional assumptions are imposed on  $T$ .

**Remark 2.10.** In Martellosio (2010), p. 162 after Equation (9) and 2nd paragraph on p. 167, it is incorrectly claimed (without providing an argument) that for any critical region of the form  $\{y \in \mathbb{R}^n : T(y) > \kappa\}$ , where  $T$  is a  $G_X$ -invariant statistic, one has

$$\text{bd}(\{y \in \mathbb{R}^n : T(y) > \kappa\}) = \text{span}(X) \cup \{y \in \mathbb{R}^n : T(y) = \kappa\}.^{16} \quad (5)$$

To see that this claim is incorrect, consider the same setting as in Example A.2 in Appendix A and let  $T = \mathbf{1}_\Phi$ . Observe that  $\Phi$  can be written as  $\{y \in \mathbb{R}^2 : T(y) > 1/2\}$  and recall that  $\Phi$  has rejection probability  $1/2$  under the null. Obviously,  $\{y \in \mathbb{R}^2 : T(y) = 1/2\} = \emptyset$  and  $\text{span}(X) = \{0\}$  hold, but

$$\text{bd}(\{y \in \mathbb{R}^2 : T(y) > 1/2\}) = \{y \in \mathbb{R}^2 : y_1 y_2 = 0\}$$

which clearly is not equal to the set  $\{0\}$ .<sup>17</sup>

Most rejection regions considered in practice (and in Martellosio (2010), see, e.g., p. 157) are of the form

$$\{y \in \mathbb{R}^n : y' C'_X B C_X y / \|C_X y\|^2 > \kappa\}, \quad (6)$$

where  $B \in \mathbb{R}^{(n-k) \times (n-k)}$  is a given symmetric matrix, which may depend on  $X$  and/or the function  $\Sigma$ , and where  $C_X$  satisfies  $C_X C'_X = I_{n-k}$  and  $C'_X C_X = \Pi_{\text{span}(X)^\perp}$ . First of all, this rejection region is strictly speaking not well-defined, as the denominator of the test statistic can take the value zero (namely, if and only if  $y \in \text{span}(X)$ ). However, whenever  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set, we then can pass to the well-defined rejection region

$$\Phi_{B,\kappa} = \Phi_{B,C_X,\kappa} = \{y \in \mathbb{R}^n : T_B(y) > \kappa\}, \quad (7)$$

where

$$T_B(y) = T_{B,C_X}(y) = \begin{cases} y' C'_X B C_X y / \|C_X y\|^2 & \text{if } y \notin \text{span}(X) \\ \lambda_1(B) & \text{if } y \in \text{span}(X), \end{cases} \quad (8)$$

without affecting the rejection probabilities. The condition that  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set is certainly satisfied if (i) the family  $\mathfrak{P}$  is absolutely continuous w.r.t. Lebesgue measure  $\mu_{\mathbb{R}^n}$  (since  $\text{span}(X)$  is a  $\mu_{\mathbb{R}^n}$ -null set in view of our assumption  $k < n$ ), or if (ii)  $\mathfrak{P}$  is an elliptically symmetric family with  $\Pr(\mathbf{z} = 0) = 0$  where  $\mathbf{z}$  is as in Assumption 3 (cf. Remark E.2(iii) in Appendix E). [Note that property (i) is always maintained in Martellosio (2010).] We shall adopt the definitions in (7) and (8) regardless of whether or not  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set. While assigning the value  $\lambda_1(B)$  to  $T_B$  on  $\text{span}(X)$  turns out to be convenient, it is of course completely arbitrary. However, assigning to  $T_B$  any other value on  $\text{span}(X)$  would, of course, have no effect on the rejection probabilities provided  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set, but it could have an effect otherwise (in which case the original definition (6) does not lead to a test at all). At any rate, an alternative assignment on  $\text{span}(X)$  has an easy to understand effect on the rejection region itself and on its boundary, see Remark 2.13 below. The test statistic  $T_B$  also depends on the choice of  $C_X$ , a dependence which is typically suppressed in the notation. Note that any other choice for  $C_X$  is necessarily of the form  $U C_X$

<sup>16</sup>This equality is trivially violated if the critical region is empty or is the entire space. However, such regions are ruled out in Martellosio (2010) as we have already noted earlier.

<sup>17</sup>Similar examples can be given when regressors are present and  $n > k + 1$ .

with  $U$  an orthogonal matrix, and thus only has the simple effect of "rotating" the matrix  $B$  as  $T_{B,C_X} = T_{UBU',UC_X}$  holds. Clearly,  $T_B$  is  $G_X$ -invariant.

Furthermore, observe that in case  $\lambda_1(B) = \lambda_{n-k}(B)$  the test statistic  $T_B$  is constant equal to  $\lambda_1(B)$ , and hence the resulting test is trivial in that the rejection region is either empty or equal to the entire sample space (depending on the choice of  $\kappa$ ). While this case is trivial in the sense that the power properties of the test are then obvious, it should be noted that this case may actually arise for commonly used tests and for certain design matrices.

The third part of the subsequent proposition now shows that for rejection regions of the form  $\Phi_{B,\kappa}$  the claim (5) regarding the boundary is indeed correct (provided  $\Phi_{B,\kappa}$  and its complement are not empty). The first part of the proposition is just a slight generalization of the observation in Martellosio (2010), Footnote 9, mentioned above. Regarding the second part we note that simple examples can be given which show that in general the inclusion can be strict (even if  $T$  is  $G_X$ -invariant).

**Proposition 2.11.** *1. If  $\Phi$  is a  $G_X$ -invariant rejection region satisfying  $\emptyset \neq \Phi \neq \mathbb{R}^n$ , then  $\text{span}(X) \subseteq \text{bd}(\Phi)$ .*

*2. If  $T$  is a test statistic which is continuous on  $\mathbb{R}^n \setminus \text{span}(X)$ , then*

$$\begin{aligned} \text{bd}(\{y \in \mathbb{R}^n : T(y) > \kappa\}) &\subseteq \text{span}(X) \cup \{y \in \mathbb{R}^n : T(y) = \kappa\} \\ &= \text{span}(X) \cup \{y \in \mathbb{R}^n \setminus \text{span}(X) : T(y) = \kappa\}. \end{aligned}$$

*3. If  $\Phi_{B,\kappa}$  is as in (7), then*

$$\begin{aligned} \text{bd}(\Phi_{B,\kappa}) &= \text{span}(X) \cup \{y \in \mathbb{R}^n : T_B(y) = \kappa\} \\ &= \text{span}(X) \cup \{y \in \mathbb{R}^n \setminus \text{span}(X) : T_B(y) = \kappa\} \end{aligned} \tag{9}$$

*provided  $\emptyset \neq \Phi_{B,\kappa} \neq \mathbb{R}^n$ .*

**Remark 2.12.** For  $\kappa < \lambda_1(B)$  we have  $\Phi_{B,\kappa} = \mathbb{R}^n$ , whereas for  $\kappa \geq \lambda_{n-k}(B)$  we have  $\Phi_{B,\kappa} = \emptyset$ . Hence, the non-trivial cases are when  $\kappa$  belongs to the interval  $[\lambda_1(B), \lambda_{n-k}(B))$  (and  $\lambda_1(B) < \lambda_{n-k}(B)$  holds). Note that in case  $\kappa = \lambda_1(B) < \lambda_{n-k}(B)$  the rejection region is the complement of a non-empty  $\mu_{\mathbb{R}^n}$ -null set (which automatically leads to the rejection probabilities being identically equal to 1 in case  $\mathfrak{P}$  is dominated by  $\mu_{\mathbb{R}^n}$ , or  $\mathfrak{P}$  is an elliptically symmetric family with  $\Pr(\mathbf{z} = 0) = 0$  where  $\mathbf{z}$  is as in Assumption 3 (cf. Remark E.2(iii) in Appendix E)), whereas for  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$  the rejection region as well as its complement have positive  $\mu_{\mathbb{R}^n}$ -measure.

**Remark 2.13.** As explained above assigning another value  $c$ , say, to  $T_B$  on  $\text{span}(X)$ , resulting in a test statistic  $T'_B$ , has no effect on the rejection probabilities provided  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set. However, it can have an effect on the resulting rejection region  $\Phi'_{B,\kappa}$ , say, and its boundary as follows: first, such a redefinition of  $T_B$  on  $\text{span}(X)$  can obviously only add  $\text{span}(X)$  to  $\Phi_{B,\kappa}$  or remove it from  $\Phi_{B,\kappa}$ . Second, inspection of the proof of Part 3 of Proposition 2.11 shows that this result continues to hold for  $\Phi'_{B,\kappa}$  provided  $\emptyset \neq \Phi'_{B,\kappa} \neq \mathbb{R}^n$  and  $\{y \in \mathbb{R}^n \setminus \text{span}(X) : T_B(y) > \kappa\} \neq \emptyset$ . In case the latter set is empty, we necessarily have  $\Phi'_{B,\kappa} = \emptyset$  or  $\Phi'_{B,\kappa} = \text{span}(X)$  (in which case Part 3 of Proposition 2.11 need not hold). But these are rather uninteresting cases as then the rejection probability is always zero provided  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set. [Also note that in these cases  $\Phi_{B,\kappa} = \emptyset$  always holds.] In particular, in the interesting case  $\kappa \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$  we have  $\Phi'_{B,\kappa} = \Phi_{B,\kappa}$  if  $c \leq \kappa$  and  $\Phi'_{B,\kappa} = \Phi_{B,\kappa} \cup \text{span}(X)$  if  $c > \kappa$ ; in both cases we have  $\text{bd}(\Phi'_{B,\kappa}) = \text{bd}(\Phi_{B,\kappa})$  and (9) also holds for  $T'_B$ .

**Correct versions of the second claim in MT1.** We next provide an assumption on the function  $\Sigma$  that will allow us to establish results which, in particular, imply a version of the second claim in MT1. The assumption may look somewhat intransparent at first sight. However, it turns out to be satisfied for commonly used correlation structures such as the ones generated by autoregressive models of order 1 or spatial autoregressions, see Sections 4.1 and 5.

**Assumption 4.** There exists a function  $c : [0, a) \rightarrow (0, \infty)$ , a normalized vector  $e \in \mathbb{R}^n$ , and a square root  $L_*(\cdot)$  of  $\Sigma(\cdot)$  such that

$$\Lambda := \lim_{\rho \rightarrow a} c(\rho) \Pi_{\text{span}(e)^\perp} L_*(\rho)$$

exists in  $\mathbb{R}^{n \times n}$  and such that the linear map  $\Lambda$  is injective when restricted to  $\text{span}(e)^\perp$ .

We note that then the image of  $\Lambda$  necessarily is  $\text{span}(e)^\perp$  and  $\Lambda$  is a bijection from  $\text{span}(e)^\perp$  to itself. As we shall see in later sections, this assumption can be verified for typical spatial models. For other types of models the equivalent condition given in the subsequent lemma is easier to verify.

**Lemma 2.14.** *Let  $c : [0, a) \rightarrow (0, \infty)$  and a normalized vector  $e \in \mathbb{R}^n$  be given. Then the function  $\Sigma(\cdot)$  satisfies Assumption 4 for the given  $c(\cdot)$ ,  $e$ , and some square root  $L_*(\cdot)$  of  $\Sigma(\cdot)$  if and only if*

$$V := \lim_{\rho \rightarrow a} c^2(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)^\perp} \quad (10)$$

*exists in  $\mathbb{R}^{n \times n}$  and the linear map  $V$  is injective when restricted to  $\text{span}(e)^\perp$ . [Necessarily the image of  $V$  is  $\text{span}(e)^\perp$  and  $V$  is a bijection from  $\text{span}(e)^\perp$  to itself.]*

**Remark 2.15.** Although Assumption 4 can hold independently of Assumption 1, the relevant case for our theory is the case where  $\Sigma$  satisfies both assumptions. If Assumptions 1 and 4 hold with  $e$  and  $e^*$ , respectively, then we claim that  $\text{span}(e) = \text{span}(e^*)$  must hold whenever  $n > 2$ . Since both conditions only depend on the span of the respective vector, we can then always choose  $e^* = e$ . To establish this claim write

$$c^2(\rho) \Pi_{\text{span}(e^*)^\perp} \Sigma(\rho) \Pi_{\text{span}(e^*)^\perp} = c^2(\rho) \lambda_n(\Sigma(\rho)) \Pi_{\text{span}(e^*)^\perp} \lambda_n^{-1}(\Sigma(\rho)) \Sigma(\rho) \Pi_{\text{span}(e^*)^\perp}$$

and note that

$$\Pi_{\text{span}(e^*)^\perp} \lambda_n^{-1}(\Sigma(\rho)) \Sigma(\rho) \Pi_{\text{span}(e^*)^\perp} \rightarrow \Pi_{\text{span}(e^*)^\perp} e e' \Pi_{\text{span}(e^*)^\perp}$$

as  $\rho \rightarrow a$  by Assumption 1. Suppose  $\text{span}(e) \neq \text{span}(e^*)$  holds. We can then find  $z \in \text{span}(e^*)^\perp$  with  $z'e \neq 0$ . But then  $z' \Pi_{\text{span}(e^*)^\perp} e e' \Pi_{\text{span}(e^*)^\perp} z = (z'e)^2 > 0$  follows. Also note that  $z' V z > 0$  where  $V$  is defined in Lemma 2.14. Together with the two preceding displays these observations imply that  $c^2(\rho) \lambda_n(\Sigma(\rho))$  converges to a finite and positive limit  $b$ , say. As a consequence,  $V = b \Pi_{\text{span}(e^*)^\perp} e e' \Pi_{\text{span}(e^*)^\perp}$  must hold, i.e.,  $V$  would have to be a matrix of rank 1. However,  $V$  is a matrix of rank  $n - 1$ , a contradiction as  $n > 2$ .

The first result is now as follows. Note that under the assumptions of the subsequent theorem the rejection probabilities actually do neither depend on  $\beta$  nor  $\sigma$ , i.e.,  $E_{\beta, \sigma, \rho}(\varphi) = E_{0, 1, \rho}(\varphi)$ , cf. Remark 2.3. For the sake of readability the subsequent two theorems are not presented in their utmost general form; possible extensions are discussed in Section 3.

**Theorem 2.16.** *Suppose Assumptions 3 and 4 hold. Let  $\varphi$  be a test that is invariant w.r.t.  $G_X$  and additionally satisfies the invariance property*

$$\varphi(y) = \varphi(y + e) \quad (11)$$

*for every  $y \in \mathbb{R}^n$  where  $e$  is as in Assumption 4. Let  $\mathcal{U}(L_*^{-1}L)$  denote the set of all accumulation points of the orthogonal matrices  $L_*^{-1}(\rho)L(\rho)$  for  $\rho \rightarrow a$ , where  $L(\rho)$  and  $L_*(\rho)$  are given in Assumptions 3 and 4, respectively. Furthermore, let  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  be arbitrary but given.*

**A.** *Suppose the distribution of  $\mathbf{z}$  (figuring in Assumption 3) possesses a density  $p$  w.r.t. Lebesgue measure  $\mu_{\mathbb{R}^n}$  that is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous. Then:*

1. *Every accumulation point of  $E_{\beta,\sigma,\rho}(\varphi)$  for  $\rho \rightarrow a$  has the form  $E_{Q_{\Lambda,U}}(\varphi)$  with  $U \in \mathcal{U}(L_*^{-1}L)$ , where  $Q_{\Lambda,U}$  denotes the distribution of  $\Lambda U \mathbf{z}$  and  $\Lambda$  is given in Assumption 4. Conversely, every element  $E_{Q_{\Lambda,U}}(\varphi)$  with  $U \in \mathcal{U}(L_*^{-1}L)$  is an accumulation point of  $E_{\beta,\sigma,\rho}(\varphi)$  for  $\rho \rightarrow a$ .*
2. *A sufficient condition for the set of accumulation points of  $E_{\beta,\sigma,\rho}(\varphi)$  for  $\rho \rightarrow a$  to be a singleton is that  $Q_{\Lambda,U}$  is the same for all  $U \in \mathcal{U}(L_*^{-1}L)$  (which, e.g., is the case if  $\mathcal{U}(L_*^{-1}L)$  is a singleton). In this case  $\lim_{\rho \rightarrow a} E_{\beta,\sigma,\rho}(\varphi)$  exists and equals  $E_{Q_{\Lambda,U}}(\varphi)$ .*
3. *Suppose the density  $p$  is such that for  $v_{S^{n-1}}$ -almost all elements  $s \in S^{n-1}$  the function  $p_s : (0, \infty) \rightarrow \mathbb{R}$  given by  $p_s(r) = p(rs)$  does not vanish  $\mu_{(0,\infty)}$ -almost everywhere. If  $\varphi$  is neither  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to zero nor  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to one, then the set of accumulation points, i.e.,  $\{E_{Q_{\Lambda,U}}(\varphi) : U \in \mathcal{U}(L_*^{-1}L)\}$ , is bounded away from zero and one.*

**B.** *Suppose  $\mathfrak{P}$  is an elliptically symmetric family with the distribution of  $\mathbf{z}$  (figuring in Assumption 3) satisfying  $\Pr(\mathbf{z} = 0) = 0$ . Then  $E_{\beta,\sigma,\rho}(\varphi)$  converges to  $E_{Q_{\Lambda,I_n}}(\varphi)$  for  $\rho \rightarrow a$  and  $E_{Q_{\Lambda,I_n}}(\varphi)$  equals  $E(\varphi(\Lambda \mathbf{G}))$  where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . Furthermore, if  $\varphi$  is neither  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to zero nor  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to one, then  $0 < E_{Q_{\Lambda,I_n}}(\varphi) < 1$  holds.*

**Remark 2.17.** (i) The condition on the density  $p$  in Part A.3 is quite weak. It is, in particular, satisfied whenever  $p$  is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set, but is much weaker. In fact, given the assumption that  $p$  exists, the condition on the density  $p$  in Part A.3 is equivalent to the assumption that the density of  $\mathbf{z}/\|\mathbf{z}\|$  w.r.t. the uniform distribution  $v_{S^{n-1}}$  on the unit sphere is  $v_{S^{n-1}}$ -almost everywhere positive; see Lemma E.1 in Appendix E. Hence, it is automatically satisfied under elliptical symmetry of  $\mathfrak{P}$  provided a density is assumed to exist.

(ii) All the conditions on the density  $p$  in Parts A.1-A.3 are certainly satisfied under the conditions used in Martellosio (2010).

(iii) Part B furthermore shows that under elliptical symmetry of  $\mathfrak{P}$  the existence of a density is in fact not required at all.

(iii) If Assumptions 3 and 4 hold with the same square root of  $\Sigma(\cdot)$  (i.e., if  $L(\cdot) = L_*(\cdot)$  can be chosen in these assumptions) as is sometimes the case, then the above theorem simplifies as  $\mathcal{U}(L_*^{-1}L)$  reduces to the singleton  $\{I_n\}$ .

(iv) Under the distributional assumptions for Part A of the preceding theorem, if  $\varphi = 0$  (or  $= 1$ )  $\mu_{\mathbb{R}^n}$ -almost everywhere then trivially  $E_{\beta,\sigma,\rho}(\varphi) = 0$  (or  $= 1$ ) holds for all  $\beta$ ,  $\sigma$ , and  $\rho$ , and hence the same holds a fortiori for the accumulation points. That the same is true under the distributional

assumptions for Part B can be seen as follows: By  $G_X$ -invariance of  $\varphi$  and the assumptions for Part B we have that  $E_{\beta,\sigma,\rho}(\varphi) = E(\varphi(L(\rho)\mathbf{G}))$  where  $\mathbf{G}$  is standard multivariate normal and  $L(\rho)$  is nonsingular, cf. (41) in Appendix C. But then  $E_{\beta,\sigma,\rho}(\varphi) = 0$  (or  $= 1$ ) follows (and the same is then a fortiori true for the limits).

(v) Similar as in Remark 2.8(ii) we make the trivial but sometimes useful observation that the limiting power of a test  $\varphi^*$  which does not satisfy the assumptions of Theorem 2.16 can nevertheless be computed from that theorem in an indirect way, if one can find another test  $\varphi$  that satisfies the assumptions of the theorem and differs from  $\varphi^*$  only on a  $\mathfrak{P}$ -null set. This remark obviously applies also to all other results in the paper and will not be repeated.

(vi) For ways of extending the results in Part B of the preceding theorem to the case where  $\Pr(\mathbf{z} = 0) > 0$  see Remark 3.1(vi) in Section 3. In a similar way Theorem 2.18 and several other results given further below can be extended to this case. We shall not mention this again.

The relationship of the preceding theorem to the second claim in MT1 is now as follows: The additional invariance property (11) in the theorem is automatically satisfied if  $e \in \text{span}(X)$  (by  $G_X$ -invariance of  $\varphi$ ). Furthermore, under the assumptions for MT1 and if  $n > 2$  the vector  $e$  in the preceding theorem coincides with  $f_1(\Sigma^{-1}(a-))$  considered in Martellosio (2010), cf. Lemma 2.5 and Remark 2.15. Hence, under Assumption 1 (which is weaker than the corresponding assumption in MT1) and if  $n > 2$ , the preceding theorem provides a substitute for the (incorrect) second claim in MT1 for the case where  $e \in \text{span}(X)$  if we specialize to  $\varphi = \mathbf{1}_\Phi$ . Recall from Proposition 2.11 that  $\text{span}(X)$  forms a part of  $\text{bd}(\Phi)$  for  $G_X$ -invariant rejection regions  $\Phi$  satisfying  $\emptyset \neq \Phi \neq \mathbb{R}^n$ . We furthermore note that the preceding theorem does not only deliver a qualitative statement like that the limiting power is strictly between 0 and 1, but provides an explicit formula for the limiting power (or the set of accumulation points). We also point out that Theorem 2.11 in Mynbaev (2012) provides related, but only qualitative, results for a certain class of rejection regions.

As just discussed, the preceding theorem provides a substitute for the second claim in MT1 in case  $e$  belongs to that part of  $\text{bd}(\Phi)$  which is represented by  $\text{span}(X)$ . If  $e \in \text{bd}(\Phi) \setminus \text{span}(X)$  then, for rejection regions  $\Phi$  of the form  $\{y \in \mathbb{R}^n : T(y) > \kappa\}$  with  $T$  satisfying a mild continuity property, Part 2 of Proposition 2.11 shows that  $\kappa = T(e)$  must hold. [Part 3 of the same proposition even shows that for the frequently used rejection regions  $\Phi_{B,\kappa}$  the conditions  $T_B(e) = \kappa$  and  $e \notin \text{span}(X)$  conversely imply  $e \in \text{bd}(\Phi_{B,\kappa}) \setminus \text{span}(X)$  provided  $\emptyset \neq \Phi_{B,\kappa} \neq \mathbb{R}^n$ .] Hence, if we can determine the limiting behavior of  $P_{\beta,\sigma,\rho}(\{y \in \mathbb{R}^n : T(y) > \kappa\})$  as  $\rho \rightarrow a$  for the case where  $\kappa = T(e)$ , this can then be used to obtain a substitute for the second claim in MT1 in case  $e \in \text{bd}(\Phi) \setminus \text{span}(X)$ , see the discussion following the subsequent theorem. This theorem now provides such a limiting result.<sup>18</sup> Like in the preceding theorem the rejection probabilities actually do neither depend on  $\beta$  nor  $\sigma$ .

**Theorem 2.18.** *Suppose Assumptions 1 and 4 hold with the same vector  $e$ , and Assumption 3 holds. Let  $T$  be a test statistic that is invariant w.r.t.  $G_X$ . Suppose there exists a positive integer  $q$  and a homogeneous multivariate polynomial  $D : \mathbb{R}^n \rightarrow \mathbb{R}$  of degree  $q$ , which does not vanish on all of  $\text{span}(e)^\perp$ , such that for every  $h \in \mathbb{R}^n$*

$$T(e+h) = T(e) + D(h) + R(h) \quad (12)$$

*holds where  $R(h)/\|h\|^q \rightarrow 0$  as  $h \rightarrow 0$ ,  $h \neq 0$ . Let  $\mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  denote the set of all accumulation points of  $(L_*^{-1}(\rho)L(\rho), \Sigma^{-1/2}(\rho)L(\rho))$  for  $\rho \rightarrow a$ . Furthermore, let  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$*

---

<sup>18</sup>It is worth noting that the assumptions of this theorem per se do not imply the assumption of Part 2 or Part 3 of Proposition 2.11.



be arbitrary but given.

1. Suppose the distribution of  $\mathbf{z}$  (figuring in Assumption 3) possesses a density  $p$  w.r.t. Lebesgue measure  $\mu_{\mathbb{R}^n}$ . Then the accumulation points of

$$P_{\beta, \sigma, \rho}(\{y \in \mathbb{R}^n : T(y) > T(e)\}) \quad (13)$$

for  $\rho \rightarrow a$  are, in case  $q$  is even, precisely given by

$$\Pr(D(\Lambda U \mathbf{z}) > 0) \quad (14)$$

with  $U \in \mathcal{U}(L_*^{-1}L)$  and where  $\Lambda$  is as in Assumption 4; for  $q$  odd, they are precisely given by

$$\Pr(D(\Lambda U \mathbf{z}) > 0, e'U_0 \mathbf{z} > 0) + \Pr(D(\Lambda U \mathbf{z}) < 0, e'U_0 \mathbf{z} < 0) \quad (15)$$

with  $(U, U_0) \in \mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$ . Thus a sufficient condition for the limit of (13) for  $\rho \rightarrow a$  to exist for even  $q$  is that  $\mathcal{U}(L_*^{-1}L)$  is a singleton, whereas for odd  $q$  it is that  $\mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  is a singleton.

2. Suppose  $\mathfrak{P}$  is an elliptically symmetric family with the distribution of  $\mathbf{z}$  satisfying  $\Pr(\mathbf{z} = 0) = 0$ . Then, if  $q$  is even,

$$\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\{y \in \mathbb{R}^n : T(y) > T(e)\}) = \Pr(D(\Lambda \mathbf{z}) > 0) = \Pr(D(\Lambda \mathbf{G}) > 0) \quad (16)$$

holds where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . If  $q$  is odd, the accumulation points of  $P_{\beta, \sigma, \rho}(\{y \in \mathbb{R}^n : T(y) > T(e)\})$  for  $\rho \rightarrow a$  are precisely given by

$$\begin{aligned} & \Pr(D(\Lambda \mathbf{z}) > 0, e'U_0 \mathbf{z} > 0) + \Pr(D(\Lambda \mathbf{z}) < 0, e'U_0 \mathbf{z} < 0) \\ &= \Pr(D(\Lambda \mathbf{G}) > 0, e'U_0 \mathbf{G} > 0) + \Pr(D(\Lambda \mathbf{G}) < 0, e'U_0 \mathbf{G} < 0) \end{aligned} \quad (17)$$

with  $U_0 \in \mathcal{U}(\Sigma^{-1/2}L_*)$  (and hence the limit of the rejection probabilities for  $\rho \rightarrow a$  necessarily exists if  $\mathcal{U}(\Sigma^{-1/2}L_*)$  is a singleton). If  $\Lambda U_0' e = 0$  holds for some  $U_0 \in \mathcal{U}(\Sigma^{-1/2}L_*)$ , the expression in (17) with this  $U_0$  then equals  $1/2$ . [A sufficient condition for  $\Lambda U_0' e = 0$  to hold is that  $\Lambda U_0'$  is symmetric.]

3. Suppose  $\mathfrak{P}$  is an elliptically symmetric family with the distribution of  $\mathbf{z}$  satisfying  $\Pr(\mathbf{z} = 0) = 0$ . If  $q$  is odd and if, additionally,

$$\lim_{\rho \rightarrow a} \lambda_n^{-1/2}(\Sigma(\rho)) c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)} = 0 \quad (18)$$

holds, where  $c(\rho)$  is as in Assumption 4, then

$$\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\{y \in \mathbb{R}^n : T(y) > T(e)\}) = 1/2.$$

We recall from Remark 2.15 that assuming that the vector  $e$  is the same in Assumptions 1 and 4 entails no loss of generality provided  $n > 2$ . Condition (18) ensures that  $\Lambda U_0' e = 0$  holds for every  $U_0 \in \mathcal{U}(\Sigma^{-1/2}L_*)$ , cf. Lemma C.2 in Appendix C, which can also be used to formulate conditions

equivalent to (18). This can be useful if one of these equivalent formulations is easier to verify in a particular application. It will turn out that condition (18) holds for autoregressive models of order 1 and certain classes of spatial error models, see Sections 4.1 and 5. Furthermore note that under the assumption that  $\mathfrak{P}$  is an elliptically symmetric family the existence of a density is not required in the preceding theorem.

Observe that, under the assumptions of MT1, the vector  $e$  in the preceding theorem coincides with  $f_1(\Sigma^{-1}(a-))$  considered in Martellosio (2010), cf. Lemma 2.5. Hence, for rejection regions of the form  $\{y \in \mathbb{R}^n : T(y) > \kappa\}$  with  $T$  satisfying the assumptions of Proposition 2.11 as well as of the preceding theorem, this theorem provides a substitute for the (incorrect) second claim in MT1 in case  $e \in \text{bd}(\Phi) \setminus \text{span}(X)$  in that it determines the limit (or the set of accumulation points) of the power function as  $\rho \rightarrow a$ . Note that the theorem itself does not in general make a statement about the limiting expressions always being strictly between 0 and 1; however, given the explicit expressions for the accumulation points of the power function in the preceding theorem, this can then be decided on a case by case basis. [We note that cases exist where the above theorem applies and the limiting power is zero or one, see, e.g., Corollary 2.23, Part 1, in case  $\lambda = \lambda_1(B)$ .]

**Remark 2.19.** (*Comments on the assumption on  $T$* ) (i) We note that under the assumptions of Theorem 2.18 the polynomial  $D$  in (12) necessarily vanishes everywhere on  $\text{span}(e)$ . More generally,  $D(h) = D(\Pi_{\text{span}(e)^\perp} h)$  holds for every  $h \in \mathbb{R}^n$ ; see Lemma C.3 in Appendix C.

(ii) If  $T$  is a test statistic that is totally differentiable at  $e$ , it satisfies relation (12) with  $q = 1$  and  $D(h) = d'h$ ,  $d$  a  $n \times 1$  vector. If  $d \notin \text{span}(e)$ , then  $D$  satisfies the assumption of the theorem. In case  $d \in \text{span}(e)$  this is not so, since  $D$  then vanishes identically on  $\text{span}(e)^\perp$  (in fact,  $d = 0$  must then hold provided  $T$  is  $G_X$ -invariant). In this case one can try to resort to higher order Taylor expansions: For example, if  $T$  is twice continuously partially differentiable in a neighborhood of  $e$ , then  $D$  can be chosen as  $1/2$  times the quadratic form corresponding to the Hessian matrix of  $T$  at the point  $e$ , provided that  $D$  does not vanish identically on  $\text{span}(e)^\perp$ .

(iii) In Theorem 2.18 the element  $e$  does not belong to the rejection region by construction. In case  $q$  is odd,  $e$  always belongs to the boundary of that region in view of homogeneity of  $D$ . The same is true in case  $q$  is even provided  $D(h) > 0$  holds for some  $h \in \mathbb{R}^n$  (which by (i) above is equivalent to  $D(h) > 0$  for some  $h \in \text{span}(e)^\perp$  with  $h \neq 0$ ). If  $q$  is even and  $D(h) < 0$  holds for all  $h \notin \text{span}(e)$  (which by (i) above is equivalent to  $D(h) < 0$  for all  $h \in \text{span}(e)^\perp$  with  $h \neq 0$ ), Lemma C.3 in Appendix C shows that then  $e$  is not an element of the boundary, but is an element of the exterior (i.e., of the complement of the closure) of the rejection region. In the remaining case, i.e.,  $q$  even and  $D(h) \leq 0$  for all  $h \notin \text{span}(e)$  but  $D(h) = 0$  for some  $h \notin \text{span}(e)$ , no conclusion can be drawn in general.

In the following example we illustrate how the assumptions on  $T$  in the preceding theorem can be verified for the important class of test statistics  $T_B$ .

**Example 2.4.** We consider the test statistic  $T_B = T_{B, C_X}$  given by (8). We assume that  $B$  is not a multiple of  $I_{n-k}$ , since otherwise  $T_B$  is constant which is a trivial case. If  $e \in \text{span}(X)$  holds, then  $T_B$  is not even continuous at  $e$ , showing that condition (12) can not be satisfied. We hence assume  $e \notin \text{span}(X)$ . Elementary calculations show that then (12) with  $q = 1$  and

$$D(h) = 2 \|C_X e\|^{-2} \left( e' C_X' B C_X - \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X' C_X \right) h \quad (19)$$

holds. In view of  $D(e) = 0$  and surjectivity of  $C_X$  we see that  $D$  does not vanish on all of  $\text{span}(e)^\perp$  if and only if  $e' C_X' B \neq \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X'$ , or in other words if and only if  $C_X e$  is not

an eigenvector of  $B$ , a condition that can easily be checked. As a point of interest we note that in this case  $T_B(e) \in (\lambda_1(B), \lambda_{n-k}(B))$  must hold, entailing that  $\emptyset \neq \Phi_{B, T_B(e)} \neq \mathbb{R}^n$  (in fact, neither  $\Phi_{B, T_B(e)}$  nor its complement are  $\mu_{\mathbb{R}^n}$ -null sets, cf. Remark 2.12). It then follows from Proposition 2.11 that  $e$  is an element of the boundary of  $\Phi_{B, T_B(e)}$ . [This can alternatively be deduced from Remark 2.19(iii).] Next consider the case where  $C_X e$  is an eigenvector of  $B$  with eigenvalue  $\lambda$ . Applying now (12) with  $q = 2$  leads to

$$D(h) = \|C_X e\|^{-2} h' (C'_X B C_X - \lambda C'_X C_X) h \quad (20)$$

which is homogeneous of degree  $q = 2$  and which does not vanish on all of  $\text{span}(e)^\perp$  (except if  $B = \lambda I_{n-k}$ , a case we have ruled out). We note that now  $T_B(e) = \lambda \in [\lambda_1(B), \lambda_{n-k}(B)]$  must hold. Recall from Remark 2.12 that in case  $\lambda$  is not the largest eigenvalue of  $B$ , we know that  $\emptyset \neq \Phi_{B, T_B(e)} \neq \mathbb{R}^n$  (in fact, neither  $\Phi_{B, T_B(e)}$  nor its complement are  $\mu_{\mathbb{R}^n}$ -null sets if additionally  $\lambda > \lambda_1(B)$  holds, whereas  $\Phi_{B, T_B(e)}$  is the complement of a non-empty  $\mu_{\mathbb{R}^n}$ -null set if  $\lambda = \lambda_1(B)$ ). Hence Proposition 2.11 shows that  $e$  then belongs to the boundary of  $\Phi_{B, T_B(e)}$ . [Since  $D(h) > 0$  holds for some  $h \in \mathbb{R}^n$  if  $\lambda$  is not the largest eigenvalue of  $B$ , this can alternatively be deduced from Remark 2.19(iii).] In case  $\lambda$  is the largest eigenvalue of  $B$ , then  $\Phi_{B, T_B(e)}$  is empty. The last case shows that, although Theorem 2.18 is geared to the case where  $e$  belongs to the boundary of the rejection region, its assumptions do not rule out other cases. Furthermore, the case where  $C_X e$  is an eigenvector of  $B$  with eigenvalue  $\lambda$  satisfying  $\lambda = \lambda_1(B)$  shows that Theorem 2.18 also applies to cases where, although  $e$  belongs to the boundary of the rejection region, the limiting rejection probabilities are not necessarily in  $(0, 1)$ .  $\square$

**Remark 2.20.** (*Comments on the set of accumulation points*) (i) If one can choose  $L_* = \Sigma^{1/2}$  in the second part of the preceding theorem then  $\mathcal{U}(\Sigma^{-1/2} L_*)$  reduces to the singleton  $\{I_n\}$  and the statement in Part 2 simplifies accordingly. A similar remark applies to the first part of the theorem in case  $L_* = L$  and/or  $L = \Sigma^{1/2}$  can be chosen.

(ii) It is not difficult to see that the accumulation points as given in (14) and (15) depend continuously on  $U$  and  $U_0$ . [This follows from the portmanteau theorem observing that  $D(\Lambda U \mathbf{z})$  as well as  $e' U_0 \mathbf{z}$  depend continuously on  $U$  and  $U_0$ , respectively, and that both expressions are nonzero almost surely as shown in the proof of Theorem 2.18.] Since  $\mathcal{U}(L_*^{-1} L)$  as well as  $\mathcal{U}(L_*^{-1} L, \Sigma^{-1/2} L)$  are compact, the question of whether or not the set of accumulation points is bounded away from 0 (or 1, respectively) then just reduces to the question as to whether every accumulation point is larger than 0 (smaller than 1, respectively). The latter question can often easily be answered by examining the explicit expressions provided by (14) and (15). For an example see the remark immediately below.

(iii) Suppose  $q = 1$  in the second part of the theorem. Observe that then  $D(h) = d'h$  with  $d \notin \text{span}(e)$  by Remark 2.19(ii). Hence, in case  $d'\Lambda$  and  $e'U_0$  are not collinear, the accumulation point given by (17) is in the open interval  $(0, 1)$ . If  $d'\Lambda$  and  $e'U_0$  are collinear, then the accumulation point is either 0 or 1.

### 2.2.3 An illustration for tests based on $T_B$

We now illustrate the results obtained so far by applying them to tests based on the statistic  $T_B = T_{B, C_X}$  defined in (8). We note that, under regularity conditions (including appropriate distributional assumptions) and excluding degenerate cases, point-optimal invariant tests and locally best invariant tests are of this form with  $B = -(C_X \Sigma(\bar{\rho}) C'_X)^{-1}$  and  $B = C_X \dot{\Sigma}(0) C'_X$ , respectively,

with  $\dot{\Sigma}(0)$  denoting the derivative at  $\rho = 0$  (ensured to exist under the aforementioned regularity conditions), see, e.g., King and Hillier (1985).<sup>19</sup>

Recall that under the assumptions in Martellosio (2010) the vector  $e$  given by Assumption 1 corresponds to the eigenvector  $f_1(\Sigma^{-1}(a-))$  in MT1, possibly up to a sign change. For that reason we impose Assumption 1 in all of the three corollaries that follow, although this assumption would not be needed for the second one of the corollaries (but note that then  $e$  would be determined by Assumption 4 only). Furthermore, recall from Remark 2.12 that  $\emptyset \neq \Phi_{B,\kappa} \neq \mathbb{R}^n$  occurs if and only if  $\kappa \in [\lambda_1(B), \lambda_{n-k}(B))$  (the interval being non-empty if and only if  $\lambda_1(B) < \lambda_{n-k}(B)$ ). We shall in the following corollaries hence always assume that  $\kappa$  is in that range and thus shall exclude the trivial cases where  $\Phi_{B,\kappa} = \emptyset$  or  $\Phi_{B,\kappa} = \mathbb{R}^n$  from the formulation of the corollaries.

The first corollary is based on Theorem 2.7. Recall that the conditions in this corollary are weaker than the conditions used in MT1 (cf. Remark 2.8) and that sufficient conditions for the high-level Assumption 2 have been given in Proposition 2.6 (under which the rejection probabilities actually do neither depend on  $\beta$  nor  $\sigma$ ).

**Corollary 2.21.** *Suppose Assumptions 1 and 2 are satisfied. Assume that  $\kappa \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$ . Then we have:*

1.  $T_B(e) > \kappa$  (i.e.,  $e \in \text{int}(\Phi_{B,\kappa})$ ) implies  $\lim_{\rho \rightarrow a} P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) = 1$  for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$ .<sup>20</sup>
2.  $T_B(e) < \kappa$  and  $e \notin \text{span}(X)$  (i.e.,  $e \notin \text{cl}(\Phi_{B,\kappa})$ ) implies  $\lim_{\rho \rightarrow a} P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) = 0$  for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$ .

It is worth pointing out here that the second case, i.e., the zero-power trap, can occur even for point-optimal invariant or locally best invariant tests as has been documented in the literature cited in the introduction. The next two corollaries now deal with the case where  $e$  belongs to the boundary of the rejection region. They are based on Theorems 2.16 and 2.18, respectively. For simplicity of presentation we concentrate only on the case of elliptically symmetric families. We remind the reader that in the two subsequent corollaries the rejection probabilities actually neither depend on  $\beta$  nor  $\sigma$ , i.e.,  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) = P_{0,1,\rho}(\Phi_{B,\kappa})$  holds.

**Corollary 2.22.** *Suppose Assumptions 1 and 4 are satisfied with the same vector  $e$ .<sup>21</sup> Furthermore, assume that  $\mathfrak{P}$  is an elliptically symmetric family (i.e., Assumption 3 holds with a spherically distributed  $\mathbf{z}$ ) and  $\Pr(\mathbf{z} = 0) = 0$ . Assume that  $\kappa \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$ . Suppose  $e \in \text{span}(X)$  holds. Then  $\lim_{\rho \rightarrow a} P_{\beta,\sigma,\rho}(\Phi_{B,\kappa})$  exists and equals  $\Pr(T_B(\Lambda \mathbf{G}) > \kappa)$  where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . Furthermore, the limit satisfies*

$$0 < \lim_{\rho \rightarrow a} P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) < 1$$

*provided  $\kappa > \lambda_1(B)$ , whereas it equals 1 in case  $\kappa = \lambda_1(B)$ .<sup>22</sup>*

<sup>19</sup>These tests are point-optimal (locally best) in the class of all  $G_X^+$ -invariant tests. As they are also  $G_X$ -invariant, they are a fortiori also point-optimal (locally best) tests in the class of  $G_X$ -invariant tests.

<sup>20</sup>Note that  $T_B(e) > \kappa$  entails  $e \notin \text{span}(X)$  in view of (8) and  $\kappa \geq \lambda_1(B)$ .

<sup>21</sup>For reasons of conformity we have here included the condition that the vector  $e$  is the same in both assumptions, although this does not impose a restriction here. This is so because of Remark 2.15 and since  $n > 2$  must hold in this corollary: Suppose  $n = 2$  would hold. Then  $k = 1$  would follow in view of  $0 \leq k < n$  and the assumption  $e \in \text{span}(X)$ . But this would be in conflict with  $\lambda_1(B) < \lambda_{n-k}(B)$ .

<sup>22</sup>In case  $\kappa = \lambda_1(B)$  the rejection region is the complement of a  $\mu_{\mathbb{R}^n}$ -null set. As discussed in Remark 2.17(iv), we then even have  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) = 1$  for every  $\beta$ ,  $\sigma$ , and  $\rho$  (although we do not require  $\mathbf{z}$  to possess a density).

The next result covers the case where  $e \in \text{bd}(\Phi_{B,\kappa}) \setminus \text{span}(X)$ . Recall from Proposition 2.11 and Remark 2.12 that this is equivalent to  $e \notin \text{span}(X)$  and  $\kappa = T_B(e) \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$ . Note that  $T_B(e) \in [\lambda_1(B), \lambda_{n-k}(B)]$  always holds by definition of  $T_B$ .

**Corollary 2.23.** *Suppose Assumptions 1 and 4 are satisfied with the same vector  $e$ . Furthermore, assume that  $\mathfrak{P}$  is an elliptically symmetric family (i.e., Assumption 3 holds with a spherically distributed  $\mathbf{z}$ ) and  $\Pr(\mathbf{z} = 0) = 0$ . Assume  $e \in \text{bd}(\Phi_{B,\kappa}) \setminus \text{span}(X)$  (i.e.,  $e \notin \text{span}(X)$  and  $\kappa = T_B(e) \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$  hold).*

1. *Suppose  $C_X e$  is an eigenvector of  $B$  with eigenvalue  $\lambda$ , say. Then  $\lambda = T_B(e) = \kappa$  and*

$$\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\Phi_{B, \kappa}) = \Pr(\mathbf{G}' \Lambda' (C_X' B C_X - \lambda C_X' C_X) \Lambda \mathbf{G} > 0), \quad (21)$$

where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . Furthermore, the limit belongs to the open interval  $(0, 1)$  if  $\lambda > \lambda_1(B)$ , whereas it equals 1 in case  $\lambda = \lambda_1(B)$ .<sup>23</sup>

2. *Suppose  $C_X e$  is not an eigenvector of  $B$ . Then the accumulation points of  $P_{\beta, \sigma, \rho}(\Phi_{B, \kappa})$  for  $\rho \rightarrow a$  are given by*

$$\begin{aligned} & \Pr\left(\left(e' C_X' B C_X - \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X' C_X\right) \Lambda \mathbf{G} > 0, e' U_0 \mathbf{G} > 0\right) + \\ & \Pr\left(\left(e' C_X' B C_X - \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X' C_X\right) \Lambda \mathbf{G} < 0, e' U_0 \mathbf{G} < 0\right) \end{aligned} \quad (22)$$

with  $U_0 \in \mathcal{U}(\Sigma^{-1/2} L_*)$ . The expression in (22) is in the open interval  $(0, 1)$  for every  $U_0 \in \mathcal{U}(\Sigma^{-1/2} L_*)$  which has the property that  $\left(e' C_X' B C_X - \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X' C_X\right) \Lambda$  and  $e' U_0$  are not collinear.<sup>24</sup> [This non-collinearity is, in particular, the case if  $\Lambda U_0' e = 0$  holds, in which case the expression in (22) equals  $1/2$ .] Furthermore, the set of all accumulation points is bounded away from 0 and 1 provided  $\left(e' C_X' B C_X - \|C_X e\|^{-2} (e' C_X' B C_X e) e' C_X' C_X\right) \Lambda$  and  $e' U_0$  are not collinear for every  $U_0 \in \mathcal{U}(\Sigma^{-1/2} L_*)$ .

3. *Suppose  $C_X e$  is not an eigenvector of  $B$ . If, additionally,*

$$\lim_{\rho \rightarrow a} \lambda_n^{-1/2}(\Sigma(\rho)) c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)} = 0$$

holds, where  $c(\rho)$  is as in Assumption 4, then

$$\lim_{\rho \rightarrow a} P_{\beta, \sigma, \rho}(\Phi_{B, \kappa}) = 1/2.$$

In the preceding corollary we have excluded the case where  $\kappa = T_B(e) = \lambda_{n-k}(B) > \lambda_1(B)$ . While we already know that this is a trivial case as then  $\Phi_{B, \kappa}$  is empty, it is interesting to note that even in this case the proof of the above corollary, which is based on Theorem 2.18, would still go through and would deliver (21), which – as it should – would then reduce to zero since the matrix  $\Lambda' (C_X' B C_X - \lambda C_X' C_X) \Lambda$  is non-positive definite in this case.<sup>25</sup>

<sup>23</sup>Cf. Footnote 22.

<sup>24</sup>If these two vectors are collinear, then the expression in (22) is 0 or 1.

<sup>25</sup>The proof of the corollary makes use of Example 2.4 which assumes  $e \notin \text{span}(X)$ . Note that  $\kappa = T_B(e) = \lambda_{n-k}(B)$  implies  $e \notin \text{span}(X)$  if  $\lambda_1(B) < \lambda_{n-k}(B)$ , allowing one to directly extend the proof of the corollary to this case.

**Remark 2.24.** (i) Appropriate versions of Corollaries 2.21-2.23 can also be given for a test statistic  $T'_B$  that takes a value  $c \neq \lambda_1(B)$  on all of  $\text{span}(X)$  and coincides with  $T_B$  on the complement of  $\text{span}(X)$ . For example, in such a version of Corollary 2.21 one needs to add the assumption  $e \notin \text{span}(X)$  in Part 1 of that corollary, because there is then no guarantee that the condition  $T'_B(e) > \kappa$  is equivalent to  $e \in \text{int}(\Phi'_{B,\kappa})$ .

(ii) In case  $\text{span}(X)$  is a  $\mathfrak{P}$ -null set (which is, e.g., the case under the assumptions of Corollaries 2.22 and 2.23, cf Remark E.2(iii)) we have  $P_{\beta,\sigma,\rho}(\Phi'_{B,\kappa}) = P_{\beta,\sigma,\rho}(\Phi_{B,\kappa})$ . Applying the above corollaries as they stand to  $T_B$  thus immediately provides information on  $P_{\beta,\sigma,\rho}(\Phi'_{B,\kappa})$  without the need of obtaining appropriate versions of the above corollaries for  $T'_B$ .

## 2.2.4 On the relationship between the size of a test and the zero-power trap

Given a  $G_X$ -invariant test statistic  $T$ , we have seen in previous sections that the limiting power of the test with rejection region  $\Phi_\kappa := \{y \in \mathbb{R}^n : T(y) > \kappa\}$  can be zero (zero-power trap). Of course, an important question to ask is for which critical values  $\kappa$  this occurs. An (essentially) equivalent formulation is to ask for which values of the sizes of the rejection regions  $\Phi_\kappa$  the limiting power is zero; i.e., for which values of the sizes the zero-power trap arises (at least along a subsequence of values of  $\rho$ ). To this end we define

$$\alpha^*(T) = \inf \left\{ P_{0,1,0}(\Phi_\kappa) : \kappa \in \mathbb{R} \text{ and } \liminf_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_\kappa) > 0 \right\}. \quad (23)$$

We note that whenever the rejection probabilities  $P_{\beta,\sigma,\rho}(\Phi_\kappa)$  are independent of  $\beta$  and  $\sigma$ , which is often the case (e.g., under Assumption 3, see Remark 2.3), the quantity  $\alpha^*(T)$  is the infimum of the sizes of all rejection regions  $\Phi_\kappa$ , the limiting power of which does not vanish. Thus  $\alpha^*(T)$  describes the size where a phase transition occurs: for sizes above  $\alpha^*(T)$  the zero-power trap does not occur, while it occurs for sizes below  $\alpha^*(T)$  (at least along a subsequence).<sup>26</sup> We investigate properties of  $\alpha^*(T)$  in this section.

Before proceeding we note that in the more narrow context of spatial regression models Martellosio (2010) also discusses the quantity  $\alpha^*(T)$  in his Lemmata D.2 and D.3, which provide the basis for a large part of the results beyond Theorem 1 in that reference.<sup>27</sup> Unfortunately, these lemmata are inappropriately stated and the proofs contain several errors. We discuss this in detail in Appendix B.1. In the present section we provide correct versions of these two lemmata, simultaneously freeing them from the spatial context, thus making them applicable to much more general covariance structures.

The subsequent lemma can now be seen as a general version of Lemma D.2 in Martellosio (2010). It gives an expression for  $\alpha^*(T)$  and shows that – under the assumptions of the lemma – for every  $\kappa$  with  $P_{0,1,0}(\Phi_\kappa) > \alpha^*(T)$  the limiting power is not only positive but in fact equals one.

**Lemma 2.25.** *Suppose Assumptions 1 and 2 are satisfied and let  $T : \mathbb{R} \rightarrow \mathbb{R}$  be a test statistic that is invariant w.r.t.  $G_X$ . Consider the family of rejection regions*

$$\Phi_\kappa = \{y \in \mathbb{R}^n : T(y) > \kappa\}$$

<sup>26</sup>While  $\alpha > \alpha^*(T)$  implies that the zero-power trap does not occur, it may in general still be the case that the limiting power is very low.

<sup>27</sup>That  $\alpha^*(T)$  defined above is indeed equivalent to the quantity  $\alpha^*$  described in Martellosio (2010), p. 165, is discussed in Appendix B.1.

for  $\kappa \in \mathbb{R}$ . Suppose there exists a  $\delta > 0$  such that  $e \notin \text{bd}(\Phi_\kappa)$  holds for every  $0 < |\kappa - T(e)| < \delta$  where  $e$  is the vector figuring in Assumption 1. [This is, in particular, satisfied if  $e \notin \text{span}(X)$  and  $T$  is continuous on  $\mathbb{R}^n \setminus \text{span}(X)$ .] If the cumulative distribution function of  $P_{0,1,0} \circ T$  is continuous at  $T(e)$ , then

$$\alpha^*(T) = P_{0,1,0}(\Phi_{T(e)}).$$

Furthermore, if for some  $\kappa$  we have  $P_{0,1,0}(\Phi_\kappa) > \alpha^*(T)$  ( $< \alpha^*(T)$ , respectively), then  $\kappa < T(e)$  ( $\kappa > T(e)$ , respectively) and  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_\kappa) = 1$  ( $= 0$ , respectively) hold.

The next result, which is based on Lemma 2.25 above, considers the test statistic  $T_B = T_{B,C_X}$  and, in particular, characterizes situations when the zero-power trap occurs or does not occur at all significance levels. Restricted to regression models with spatial autoregressive errors of order one, the subsequent lemma contains a correct and improved version of Lemma D.3 in Martellosio (2010) as a special case, the improvement relating amongst others to the fact that we do not only characterize when  $\alpha^*(T_B)$  equals 0 or 1, but that we also determine the limiting power in each case. Before presenting the result, we note that Lemma D.3 in Martellosio (2010) is stated for tests obtained from  $T_B$  by rejecting for small values of the test statistic while we state our result for tests that reject for large values of  $T_B$ . However, this is immaterial as Lemma D.3 in Martellosio (2010) can trivially be rephrased in our setting by simply passing from  $B$  to  $-B$ . In the subsequent two propositions we exclude the trivial case where  $\lambda_1(B) = \lambda_{n-k}(B)$  holds, in which case  $\alpha^*(T_B) = 1$ . [To see this note that then  $T_B$  is constant equal to  $\lambda_1(B)$  and thus all rejection probabilities are zero or one depending on whether  $\kappa \geq \lambda_1(B)$  or  $\kappa < \lambda_1(B)$ .]

**Proposition 2.26.** *Suppose Assumptions 1 and 2 hold. Furthermore, assume that  $P_{0,1,0}$  is  $\mu_{\mathbb{R}^n}$ -absolutely continuous with a density that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set. Suppose  $e \notin \text{span}(X)$  where  $e$  is the vector figuring in Assumption 1 and suppose that  $\lambda_1(B) < \lambda_{n-k}(B)$  holds. Then:*

1.  $\alpha^*(T_B) = 0$  if and only if  $C_X e \in \text{Eig}(B, \lambda_{n-k}(B))$ . If  $C_X e \in \text{Eig}(B, \lambda_{n-k}(B))$  holds, then  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) = 1$  for every  $\kappa \in (-\infty, \lambda_{n-k}(B))$ . [For  $\kappa \geq \lambda_{n-k}(B)$  we trivially always have  $\Phi_{B,\kappa} = \emptyset$ .]
2.  $\alpha^*(T_B) = 1$  if and only if  $C_X e \in \text{Eig}(B, \lambda_1(B))$ . If  $C_X e \in \text{Eig}(B, \lambda_1(B))$  holds, then  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) = 0$  for every  $\kappa \in (\lambda_1(B), \infty)$ . [For  $\kappa < \lambda_1(B)$  we trivially always have  $\Phi_{B,\kappa} = \mathbb{R}^n$ , whereas  $\Phi_{B,\kappa}$  is the complement of a  $\mu_{\mathbb{R}^n}$ -null set in case  $\kappa = \lambda_1(B) < \lambda_{n-k}(B)$ .<sup>28</sup>]
3.  $0 < \alpha^*(T_B) < 1$  if and only if  $C_X e$  neither belongs to  $\text{Eig}(B, \lambda_1(B))$  nor  $\text{Eig}(B, \lambda_{n-k}(B))$ . If  $C_X e$  neither belongs to  $\text{Eig}(B, \lambda_1(B))$  nor  $\text{Eig}(B, \lambda_{n-k}(B))$ , there exists a unique  $\kappa^* \in (\lambda_1(B), \lambda_{n-k}(B))$  such that  $P_{0,1,0}(\Phi_{B,\kappa^*}) = \alpha^*(T_B)$ ; furthermore,  $\kappa^* = T_B(e)$  holds, and for  $\kappa < \kappa^*$  ( $\kappa > \kappa^*$ , respectively) we have  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) = 1$  ( $= 0$ , respectively).

Part 3 is silent on the limiting power in case  $\kappa = \kappa^*$ . Under additional assumptions, information on the limiting power in this case has been provided in Corollary 2.23; we do not repeat the results. Furthermore, note that in view of Lemma C.4 in Appendix C the analogon to  $\kappa^*$  in Part 1 is  $\lambda_{n-k}(B)$ , whereas in Part 2 it is  $\lambda_1(B)$ .

---

<sup>28</sup>Hence,  $P_{0,1,0}(\Phi_{B,\kappa}) = 1$  holds in case  $\kappa = \lambda_1(B) < \lambda_{n-k}(B)$ . Furthermore,  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) = 1$  will then also hold provided, e.g., the measures  $P_{0,1,\rho}$  are  $\mu_{\mathbb{R}^n}$ -absolutely continuous.

In the case of a pure correlation model, i.e.,  $k = 0$ , the condition  $e \notin \text{span}(X)$  is always satisfied and the preceding lemma tells us that the limiting power of the test based on  $T_B$  is then always 1 (for every choice of  $\kappa < \lambda_{n-k}(B)$ ), and thus the power-trap never arises, if and only if  $e$  is an eigenvector of  $B$  to the eigenvalue  $\lambda_{n-k}(B)$ .

As already noted in the discussion following Corollary 2.21, point-optimal invariant as well as locally best invariant tests are in general not guaranteed to be immune to the zero-power trap phenomenon, i.e., they can fall under the wrath of Case 2 or 3 of the preceding proposition. However, under its assumptions, Proposition 2.26 also tells us how we may construct – for a given covariance model  $\Sigma(\cdot)$  and a given design matrix  $X$  – a test that avoids the zero-power trap and even has limiting power equal to 1: All that needs to be done is to choose  $B$  such that  $C_X e \in \text{Eig}(B, \lambda_{n-k}(B))$  holds; one such choice is given by  $B = C_X e e' C_X'$ , but there are many other choices. However, this result does not tell us anything about whether or not such a test has good power properties for  $\rho$  not close to  $a$ . For more on ways to overcome the zero-power trap see Preinerstorfer (2014).

**Remark 2.27.** Suppose the rejection probabilities  $P_{\beta, \sigma, \rho}(\Phi_{B, \kappa})$  are independent of  $\beta$  and  $\sigma$  (which is, e.g., the case under Assumption 3 (see Remark 2.3)) and suppose  $P_{0,1,0}$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ . Furthermore assume that  $\lambda_1(B) < \lambda_{n-k}(B)$  holds. Then it follows from our Lemma C.4 in Appendix C that for every  $\alpha \in (0, 1)$  one can find a  $\kappa(\alpha) \in (\lambda_1(B), \lambda_{n-k}(B))$  such that  $\Phi_{B, \kappa(\alpha)}$  has size  $\alpha$ . If, additionally,  $P_{0,1,0}$  has a density that is positive on an open neighborhood of the origin except possibly for an  $\mu_{\mathbb{R}^n}$ -null set, then  $\kappa(\alpha)$  is unique and satisfies  $\kappa(\alpha) \rightarrow \lambda_1(B)$  ( $\rightarrow \lambda_{n-k}(B)$ ) as  $\alpha \rightarrow 1$  ( $\rightarrow 0$ ).

While Proposition 2.26 concerns the case  $e \notin \text{span}(X)$ , we have, as a simple consequence of Theorem 2.16, the following result in case  $e \in \text{span}(X)$ . In contrast to the cases discussed in the preceding proposition only the case  $\alpha^*(T_B) = 0$  can occur. Recall from Remark 2.15 that whenever Assumptions 1 and 4 both hold, the vector  $e$  in the subsequent proposition is the same as the vector  $e$  in Proposition 2.26 above (since  $n > 2$  must hold in the subsequent proposition). Also recall that the rejection probabilities do neither depend on  $\beta$  nor  $\sigma$  under the assumption of the subsequent proposition, hence the results could be rephrased for  $P_{\beta, \sigma, \rho}(\Phi_{B, \kappa})$  where  $\beta$  and  $\sigma$  are arbitrary.

**Proposition 2.28.** *Suppose Assumptions 3 and 4 hold. Furthermore, assume that the distribution of  $\mathbf{z}$  (figuring in Assumption 3) possesses a density  $p$  w.r.t. Lebesgue measure  $\mu_{\mathbb{R}^n}$ , which is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous and which is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set. Suppose  $e \in \text{span}(X)$ , where  $e$  is the vector figuring in Assumption 4 and suppose that  $\lambda_1(B) < \lambda_{n-k}(B)$  holds. Then  $\alpha^*(T_B) = 0$  always holds. Furthermore,*

$$0 < \liminf_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) \leq \limsup_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) < 1$$

*holds for every  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$ , whereas*

$$\liminf_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{B,\kappa}) = 1$$

*holds for  $\kappa \leq \lambda_1(B)$ . [For  $\kappa \geq \lambda_{n-k}(B)$  we trivially always have  $\Phi_{B,\kappa} = \emptyset$ .]*

**Remark 2.29.** The assumption in the preceding proposition, that  $p$  is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set, can be replaced by the weaker assumption used in Part A.3 in Theorem 2.16. Furthermore, the assumption that a density  $p$  exists can be completely removed if  $\mathfrak{P}$  is assumed to be an elliptically symmetric family with  $\Pr(\mathbf{z} = 0) = 0$ .



### 2.3 On indistinguishability by invariant tests

The discussion so far has been concerned with evaluating the power function of a  $G_X$ -invariant test for values of  $\rho$  close to  $a$ , the upper bound of the range of  $\rho$ . In particular, we have identified conditions under which the power function approaches zero for  $\rho \rightarrow a$  (zero-power trap). These conditions, of course, depend on the test considered as well as on the underlying model. In this section we now isolate conditions on the model alone under which the null and alternative hypotheses are indistinguishable by *any*  $G_X$ -invariant test (in fact, by any  $G_X^1$ -invariant test) whatsoever. These results, given in Theorem 2.30 and Corollary 2.31 below, contain a number of results in the literature as special cases: (i) The univariate case of Theorem 5 in Arnold (1979) concerning flatness of the power function of invariant tests in a linear model with intercept and exchangeably distributed errors, (ii) Theorem 5 in Kadiyala (1970), (iii) those parts of Propositions 3-5 in Martellosio (2010) regarding flatness of the power function of the tests considered there (see Section 4.3 for further discussion), (iv) the first half of the theorem proved in Martellosio (2011b) (see also Section 4.3), and (v) the result on the likelihood ratio test in Kariya (1980).

**Theorem 2.30.** *Suppose that for some  $0 < \rho^* < a$  the matrix  $C_X \Sigma(\rho^*) C_X'$  is a multiple of  $I_{n-k}$ , i.e.,  $C_X \Sigma(\rho^*) C_X' = \delta(\rho^*) I_{n-k}$ .*

1. *Then for every  $n \times n$  matrix  $K(\rho^*)$  satisfying  $K(\rho^*) K'(\rho^*) = \Sigma(\rho^*)$  there exists an orthogonal  $n \times n$  matrix  $U(\rho^*)$  such that for every  $\beta \in \mathbb{R}^k$  and every  $0 < \sigma < \infty$ ,*

$$\begin{aligned} \mathcal{I}_X(X\beta + \sigma K(\rho^*)z) &= \mathcal{I}_X(U(\rho^*)z), \\ \mathcal{I}_X^+(X\beta + \sigma K(\rho^*)z) &= \mathcal{I}_X^+(U(\rho^*)z), \\ \mathcal{I}_X^1(X\beta + \sigma K(\rho^*)z) &= \mathcal{I}_X^1(\sigma \delta^{1/2}(\rho^*) U(\rho^*)z) \end{aligned} \tag{24}$$

*hold for every  $z \in \mathbb{R}^n$ , where  $\mathcal{I}_X$ ,  $\mathcal{I}_X^+$ , and  $\mathcal{I}_X^1$  have been defined in Section 2.1.1.*

2. *Suppose, furthermore, that  $\mathfrak{P}$  is an elliptically symmetric family. Then for every  $\beta \in \mathbb{R}^k$  and every  $0 < \sigma < \infty$*

$$P_{\beta, \sigma, \rho^*} \circ \mathcal{I}_X = P_{\beta, \sigma, 0} \circ \mathcal{I}_X = P_{0, 1, 0} \circ \mathcal{I}_X,$$

*and*

$$P_{\beta, \sigma, \rho^*} \circ \mathcal{I}_X^+ = P_{\beta, \sigma, 0} \circ \mathcal{I}_X^+ = P_{0, 1, 0} \circ \mathcal{I}_X^+,$$

*whereas*

$$P_{\beta, \sigma, \rho^*} \circ \mathcal{I}_X^1 = P_{0, \sigma \delta^{1/2}(\rho^*), 0} \circ \mathcal{I}_X^1 \quad \text{and} \quad P_{\beta, \sigma, 0} \circ \mathcal{I}_X^1 = P_{0, \sigma, 0} \circ \mathcal{I}_X^1.$$

*The same relations hold with  $\mathcal{I}_X$ ,  $\mathcal{I}_X^+$ , and  $\mathcal{I}_X^1$ , respectively, replaced by arbitrary  $G_X$ -,  $G_X^+$ -, or  $G_X^1$ -invariant statistics, meaning that no  $G_X^1$ -invariant test (and a fortiori no  $G_X^+$ -invariant or  $G_X$ -invariant test) can distinguish the null  $H_0$  defined in (3) from the alternative  $\rho = \rho^*$ ,  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ . In particular, the power function of any  $G_X^+$ -invariant test (and a fortiori of any  $G_X$ -invariant test) is constant on  $\mathbb{R}^k \times (0, \infty) \times \{0, \rho^*\}$ , whereas for any  $G_X^1$ -invariant test power is always less than or equal to size.*

**Corollary 2.31.** *Suppose  $C_X \Sigma(\rho^*) C_X'$  is a multiple of  $I_{n-k}$  for every  $\rho^* \in (0, a)$  and  $\mathfrak{P}$  is an elliptically symmetric family. Then no  $G_X^1$ -invariant test (and a fortiori no  $G_X^+$ -invariant or  $G_X$ -invariant test) can distinguish  $H_0$  from the alternative  $H_1$  defined in (3). In particular, the power function of any  $G_X^+$ -invariant test (and a fortiori of any  $G_X$ -invariant test) is constant on  $\mathbb{R}^k \times (0, \infty) \times [0, a)$ , whereas for any  $G_X^1$ -invariant test power is always less than or equal to size.*

**Remark 2.32.** (i) The condition that  $C_X \Sigma(\rho^*) C'_X$  is a multiple of  $I_{n-k}$  does not depend on the particular choice of  $C_X$  as any two such choices differ only by premultiplication with an orthogonal matrix. Furthermore, note that the condition  $C_X \Sigma(\rho^*) C'_X = \delta(\rho^*) I_{n-k}$  is equivalent to  $\Pi_{\text{span}(X)^\perp} \Sigma(\rho^*) \Pi_{\text{span}(X)^\perp} = \delta(\rho^*) \Pi_{\text{span}(X)^\perp}$ , see Lemma C.5 in Appendix C.

(ii) Suppose that for some  $0 < \rho^* < a$  the matrix  $C_X \Sigma(\rho^*) C'_X$  is not a multiple of  $I_{n-k}$  and that  $\mathfrak{P}$  is an elliptically symmetric family. Then it can be shown that for every  $\alpha \in (0, 1)$  there exists a  $G_X$ -invariant size  $\alpha$  test with power at  $(\beta, \sigma, \rho^*)$  strictly larger than  $\alpha$  for every  $\beta \in \mathbb{R}^k$  and every  $0 < \sigma < \infty$ . As a consequence of this result and Theorem 2.30 we see that the hypothesis  $\rho = 0$  and the alternative  $\rho = \rho^*$  are distinguishable by a  $G_X$ -invariant ( $G_X^+$ -invariant,  $G_X^1$ -invariant) test if and only if  $C_X \Sigma(\rho^*) C'_X$  is not a multiple of  $I_{n-k}$ . [If Assumption 3 is satisfied but  $\mathfrak{P}$  is not an elliptically symmetric family, the hypothesis  $\rho = 0$  and the alternative  $\rho = \rho^*$  may still be distinguishable by a  $G_X$ -invariant test even in the case where  $C_X \Sigma(\rho^*) C'_X$  is a multiple of  $I_{n-k}$ , provided  $L(\rho^*)$  from Assumption 3 gives rise to a  $U(\rho^*) \neq L(0)$  when it is used for  $K(\rho^*)$  in Part 1 of Theorem 2.30.]

**Remark 2.33.** (*Generalization of Theorem 2.30 and Corollary 2.31*) Part 2 of Theorem 2.30 is true more generally if  $\mathfrak{P}$  satisfies Assumption 3 and if  $\Pi_{\text{span}(X)^\perp} L(\rho^*) \mathbf{z}$  has the same distribution as a positive multiple of  $\Pi_{\text{span}(X)^\perp} L(0) \mathbf{z}$ , where  $L(\cdot)$  and  $\mathbf{z}$  are as in Assumption 3 (the multiple then being necessarily equal to  $\delta^{1/2}(\rho^*)$ ). A sufficient condition for this clearly is that  $\Pi_{\text{span}(X)^\perp} L(\rho^*)$  is a positive multiple of  $\Pi_{\text{span}(X)^\perp} L(0)$ , for which in turn a sufficient condition is that both of these two matrices are a multiple of  $\Pi_{\text{span}(X)^\perp}$  with the multiples being non-zero and having the same sign. Similarly, Corollary 2.31 holds if  $\mathfrak{P}$  satisfies Assumption 3 and the distributions of  $\delta^{-1/2}(\rho^*) \Pi_{\text{span}(X)^\perp} L(\rho^*) \mathbf{z}$  for  $\rho^* \in [0, a)$  do not depend on  $\rho^*$  (a sufficient condition for this being that  $\Pi_{\text{span}(X)^\perp} L(\rho^*)$  is a positive multiple of  $\Pi_{\text{span}(X)^\perp} L(0)$  for every  $\rho^* \in [0, a)$ ). Such cases arise naturally in the context of spatial models, see Section 4.3.

**Remark 2.34.** Suppose Assumption 1 holds and  $C_X \Sigma(\rho^*) C'_X = \delta(\rho^*) I_{n-k}$  for all  $\rho^* \in (0, a)$  (or at least for a sequence  $\rho_m^*$  converging to  $a$ ). It is then not difficult to see that then either  $e \in \text{span}(X)$  or  $n = k + 1$  must hold.<sup>29</sup>

Theorem 2.30 explains the flatness of power functions of  $G_X$ - (or  $G_X^+$ -) invariant tests observed in the literature cited above in terms of an identification problem in the "reduced" experiment, where the reduction is effected by the action of the group  $G_X$  (or  $G_X^+$ ) (i.e., the parameters are not identifiable from the distribution of the corresponding maximal invariant statistic); cf. Remark 2 in Martellosio (2011b) for a special case. In our framework this shows that what has been dubbed *non-identifiability as a hypothesis* in Kariya (1980) is simply an identification problem in the distribution of the maximal invariant statistic.

### 3 Some generalizations

**Remark 3.1.** (*Generalizations of the distributional assumptions*) (i) We start with the following simple observation: Suppose Assumption 3 holds with  $\Pr(\mathbf{z} = 0) = 0$ . Let  $\mathbf{z}^\dagger$  be another random vector of the same dimension as  $\mathbf{z}$  (possibly defined on another probability space) with  $\Pr(\mathbf{z}^\dagger = 0) = 0$  and such that  $\mathbf{z}^\dagger / \|\mathbf{z}^\dagger\|$  has the same distribution as  $\mathbf{z} / \|\mathbf{z}\|$ . It is then easy to see that the rejection probabilities of any  $G_X$ -invariant (or  $G_X^+$ -invariant) test are the same whether they are computed

---

<sup>29</sup>Cf. Footnote 39.

under  $P_{\beta,\sigma,\rho}$  or under  $P_{\beta,\sigma,\rho}^\dagger$ , where  $P_{\beta,\sigma,\rho}^\dagger$  is the distribution of  $\mathbf{y}^\dagger$  which is obtained from  $\mathbf{z}^\dagger$  via Assumption 3 in the same way as  $\mathbf{y}$  is obtained from  $\mathbf{z}$ . Hence, any result that holds for rejection probabilities of a  $G_X$ -invariant (or  $G_X^+$ -invariant) test obtained under model  $\mathfrak{P}^\dagger$  automatically carries over to the rejection probabilities of the same test obtained under model  $\mathfrak{P}$ .

(ii) An immediate consequence of the preceding observation is, for example, that Part 1 of Theorem 2.18 continues to hold if the requirement that  $\mathbf{z}$  has a density is replaced by the following weaker condition (just apply Part 1 of Theorem 2.18 to  $\mathfrak{P}^\dagger$ ):

*Condition (\*):*  $\Pr(\mathbf{z} = 0) = 0$  and there exists a random vector  $\mathbf{z}^\dagger$ , which possesses a density  $p^\dagger$  w.r.t. Lebesgue measure, such that  $\mathbf{z}/\|\mathbf{z}\|$  and  $\mathbf{z}^\dagger/\|\mathbf{z}^\dagger\|$  have the same distribution.

This condition can be shown to be equivalent to the more explicit condition that  $\Pr(\mathbf{z} = 0) = 0$  and that  $\mathbf{z}/\|\mathbf{z}\|$  possesses a density with respect to the uniform probability measure  $v_{S^{n-1}}$  on  $S^{n-1}$ , see Lemmata E.1 and E.3 in Appendix E. As a consequence, Part 1 of Theorem 2.18 could have been stated more generally under the assumption that  $\Pr(\mathbf{z} = 0) = 0$  and that  $\mathbf{z}/\|\mathbf{z}\|$  possesses a density with respect to the uniform probability measure  $v_{S^{n-1}}$  on  $S^{n-1}$ .

(iii) The same reasoning as in (ii) shows that Part A of Theorem 2.16 holds even without the assumption of absolute continuity of the distribution of  $\mathbf{z}$  under the following weaker assumptions: Parts A.1 and A.2 hold provided Condition (\*) is satisfied and provided  $\mathbf{z}^\dagger$  can be chosen in such a way that the density  $p^\dagger$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous; an explicit sufficient condition for this is that  $\Pr(\mathbf{z} = 0) = 0$  holds and that  $\mathbf{z}/\|\mathbf{z}\|$  possesses a  $v_{S^{n-1}}$ -almost everywhere continuous density, see Lemma E.3 in Appendix E. [Unfortunately, this explicit condition is not necessary, making it difficult to give a simple equivalent condition which is in terms of the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  only.] Furthermore, Part A.3 holds, provided Condition (\*) is satisfied and provided  $\mathbf{z}^\dagger$  can be chosen in such a way that the density  $p^\dagger$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous and has the property that for  $v_{S^{n-1}}$ -almost all  $s \in S^{n-1}$  the function  $p^\dagger(rs)$  does not vanish  $\mu_{(0,\infty)}$ -almost everywhere. An explicit sufficient condition for this is that  $\Pr(\mathbf{z} = 0) = 0$  and  $\mathbf{z}/\|\mathbf{z}\|$  possesses a  $v_{S^{n-1}}$ -almost everywhere continuous and  $v_{S^{n-1}}$ -almost everywhere positive density, see Lemmata E.1 and E.3 in Appendix E.

(iv) In case  $\mathfrak{P}$  is an elliptically symmetric family with  $\Pr(\mathbf{z} = 0) = 0$  then  $\mathbf{z}$  is spherically symmetric entailing that the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  is the uniform distribution on the unit sphere  $S^{n-1}$ . Hence, the explicit conditions discussed above are met, entailing that in this case Condition (\*) is always satisfied and  $\mathbf{z}^\dagger$  can be chosen such that  $p^\dagger$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous and  $\mu_{\mathbb{R}^n}$ -almost everywhere positive (in fact,  $\mathbf{z}^\dagger$  can be chosen to be Gaussian). This is what underlies Parts 2 and 3 of Theorem 2.18 as well as Part B of Theorem 2.16.

(v) Suppose  $\mathfrak{P}$  does not satisfy Assumption 3 but each element  $P_{\beta,\sigma,\rho}$  of  $\mathfrak{P}$  is elliptically symmetric and does not have an atom at  $X\beta$  (that is,  $\mathbf{u}$  is now distributed as  $\sigma L(\rho)\mathbf{w}$  where  $\mathbf{w}$  has zero mean, identity covariance matrix, and is spherically symmetric with  $\Pr(\mathbf{w} = 0) = 0$ , but where the distribution of  $\mathbf{w}$  now may depend on the parameters  $\beta, \sigma, \rho$ ). Then it follows from the results in Appendix E and from the argument underlying the discussion in (i) above that we may replace  $\mathfrak{P}$  by an *elliptically symmetric family*  $\mathfrak{P}^\dagger$  (even by a Gaussian family) without affecting the rejection probabilities of  $G_X$ -invariant (or  $G_X^+$ -invariant) tests and then apply our results. [More generally, if  $\mathbf{w}$  is not necessarily spherically symmetric, but the distribution of  $\mathbf{w}/\|\mathbf{w}\|$  does not depend on the parameters  $\beta, \sigma, \rho$ , we may replace  $\mathfrak{P}$  by a family  $\mathfrak{P}^\dagger$  that is based on a  $\mathbf{z}^\dagger$ , the distribution of which does not depend on the parameters, and consequently satisfies Assumption 3.]

(vi) In the above discussion we have so far not considered cases where Assumption 3 holds, but  $\vartheta := \Pr(\mathbf{z} = 0)$  is positive. These cases can be treated as follows: Observe that then  $P_{\beta,\sigma,\rho} = \vartheta\delta_{X\beta} + (1 - \vartheta)\tilde{P}_{\beta,\sigma,\rho}$  where now  $\tilde{P}_{\beta,\sigma,\rho}$  satisfies Assumption 3 and the corresponding  $\tilde{\mathbf{z}}$  has no

mass at the origin (and is spherically symmetric if  $\mathbf{z}$  is so). Now for a  $G_X$ -invariant ( $G_X^+$ -invariant,  $G_X^1$ -invariant) test  $\varphi$  the rejection probabilities satisfy  $E_{\beta,\sigma,\rho}(\varphi) = \vartheta\varphi(X\beta) + (1 - \vartheta)\tilde{E}_{\beta,\sigma,\rho}(\varphi)$ , where we observe that  $\varphi(X\beta) = \varphi(0)$  is a constant not depending on  $\beta$  (due to invariance of  $\varphi$ ). Hence, the behavior of  $E_{\beta,\sigma,\rho}(\varphi)$  can be deduced from the behavior of  $\tilde{E}_{\beta,\sigma,\rho}(\varphi)$ , to which our results are applicable.

**Remark 3.2.** (*Semiparametric Models*) Throughout the paper we have taken a parametric viewpoint in that the distribution of  $\mathbf{y}$  is assumed to be completely determined by the parameters  $\beta$ ,  $\sigma$ , and  $\rho$ . The above discussion shows that some of the results of the paper like Theorems 2.16 and 2.18 also apply in broader semiparametric settings (as only properties of the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  and  $\Pr(\mathbf{z} = 0) = 0$  matter). To give just one example, let  $\mathfrak{P}_{all}$  denote the *semiparametric* model of *all* elliptically symmetric distributions with mean  $X\beta$  and covariance matrix  $\sigma^2\Sigma(\rho)$  that have no atom at  $X\beta$  and where  $(\beta, \sigma, \rho)$  varies in  $\mathbb{R}^k \times (0, \infty) \times [0, a)$ . The preceding discussion then shows that the rejection probabilities of a  $G_X$ -invariant test coincide with the rejection probabilities of a corresponding parametric elliptically symmetric family  $\mathfrak{P}$  (which actually can be assumed to be Gaussian). Hence, the behavior of the rejection probabilities corresponding to  $\mathfrak{P}_{all}$  can immediately be deduced from Theorems 2.16 and 2.18 (applied to  $\mathfrak{P}$ ).

**Remark 3.3.** (*Extensions to  $G_X^+$ -invariant tests*) The results of the present paper, apart from a few exceptions, are concerned with properties of  $G_X$ -invariant tests. Concentrating on  $G_X$ -invariant tests, however, does not seem to impose a serious restriction since most tests for the testing problem (3) available in the literature satisfy this invariance property. If one nevertheless is interested in the larger class of  $G_X^+$ -invariant tests, the following observation is of interest as it allows one to extend our results to this larger class of tests: Suppose Assumption 3 holds with the vector  $\mathbf{z}$  having the same distribution as  $-\mathbf{z}$  (which, in particular, is the case under spherical symmetry). For a  $G_X^+$ -invariant test  $\varphi$  define the test  $\varphi^*$  by  $\varphi^*(y) = (\varphi(y) + \varphi(-y))/2$ , which clearly is  $G_X$ -invariant. Furthermore,  $E_{\beta,\sigma,\rho}\varphi = E_{\beta,\sigma,\rho}\varphi^*$  holds for every  $\beta$ ,  $\sigma$ , and  $\rho$ . Applying now our results to  $\varphi^*$  then delivers corresponding results for  $\varphi$ .

**Remark 3.4.** (*Further Generalizations*) Our results easily extend to the case where the covariance model is defined only on a set  $R$ , with  $0 \in R \subseteq [0, a)$ , that has  $a$  as its accumulation point. This observation, in particular, allows one to obtain limiting power results along certain sequences  $\rho_m$ ,  $\rho_m \rightarrow a$ , when some of the assumptions (like Assumptions 1, 2, 3, or 4) hold only along these sequences.

## 4 An application to spatial regression models

In this section we focus on regression models with spatial autoregressive errors of order one, i.e., SAR(1) disturbances, and on spatial lag models. First, we consider the case of a regression model with SAR(1) errors, i.e., what is sometimes also called a spatial error model. Second, we consider a spatial lag model.

### 4.1 Spatial error models

Let  $n \geq 2$  and let  $W$  be a given  $n \times n$  matrix, the *weights matrix*. We assume that the diagonal elements of  $W$  are all zero and that  $W$  has a positive (real) eigenvalue, denoted by  $\lambda_{\max}$ , such that any other real or complex zero of the characteristic polynomial of  $W$  is in absolute value not

larger than  $\lambda_{\max}$ . We also assume that  $\lambda_{\max}$  has algebraic multiplicity (and thus also geometric multiplicity) equal to 1. Choose  $f_{\max}$  as a normalized eigenvector of  $W$  corresponding to  $\lambda_{\max}$  (which is unique up to multiplication by  $-1$ ). The spatial error model (SEM) is then given by the regression model in equation (1) where the disturbance vector  $\mathbf{u}$  is SAR(1), i.e., for given  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and  $\rho \in [0, \lambda_{\max}^{-1})$  we have

$$\mathbf{u} = \rho W \mathbf{u} + \sigma \boldsymbol{\varepsilon} \quad (25)$$

where  $\boldsymbol{\varepsilon}$  is a mean zero random vector with covariance matrix  $I_n$ . Observe that then clearly

$$\mathbf{u} = (I_n - \rho W)^{-1} \sigma \boldsymbol{\varepsilon} \quad (26)$$

holds and that the covariance matrix of  $\mathbf{u}$  is given by  $\sigma^2 \Sigma_{SEM}(\rho)$  where  $\Sigma_{SEM}(\rho) = [(I_n - \rho W')(I_n - \rho W)]^{-1}$  for  $\rho \in [0, a)$  where here  $a = \lambda_{\max}^{-1}$ . Additionally we assume that the distribution of  $\boldsymbol{\varepsilon}$  is a fixed distribution independent of  $\beta$ ,  $\sigma$ , and  $\rho$ .<sup>30</sup> *The above are the maintained assumptions for the SEM considered in this section.* The parametric family  $\mathfrak{P}$  of probability measures induced by (1) and (25) under the maintained assumptions will be denoted by  $\mathfrak{P}_{SEM}$ .

**Remark 4.1.** If  $W$  is an (elementwise) nonnegative and irreducible matrix with zero elements on the main diagonal, a frequent assumption for spatial weights matrices, then the above assumptions on  $W$  are satisfied by the Perron-Frobenius theorem and  $\lambda_{\max}$  is then the Perron-Frobenius root of  $W$  (see, e.g., Horn and Johnson (1985), Theorem 8.4.4, p. 508). In this case one can always choose  $f_{\max}$  to be entrywise positive.

The next lemma shows identifiability of the parameters in the model, identifiability of  $\beta$  being trivial. An immediate consequence is that the two subsets of  $\mathfrak{P}_{SEM}$  corresponding to the null hypothesis  $\rho = 0$  and alternative hypothesis  $\rho > 0$  are disjoint.<sup>31</sup>

**Lemma 4.2.** *If  $\sigma_1^2 \Sigma_{SEM}(\rho_1) = \sigma_2^2 \Sigma_{SEM}(\rho_2)$  holds for  $\rho_i \in [0, \lambda_{\max}^{-1})$  and  $0 < \sigma_i < \infty$  ( $i = 1, 2$ ) then  $\rho_1 = \rho_2$  and  $\sigma_1 = \sigma_2$ .*

We next verify that the spatial error model satisfies Assumptions 1, 3, and 4, and that it satisfies Assumption 2 under a mild condition on the distribution of  $\boldsymbol{\varepsilon}$ . The first claim in Lemma 4.3 also appears in Martellosio (2011b), Lemma 3.3.

**Lemma 4.3.**  *$\Sigma_{SEM}(\cdot)$  satisfies Assumption 1 with  $e = f_{\max}$  as well as Assumption 4 with  $e = f_{\max}$ ,  $c(\rho) = 1$ ,  $L_*(\rho) = (I_n - \rho W)^{-1}$ , and  $\Lambda = (I_n - \lambda_{\max}^{-1} \Pi_{\text{span}(f_{\max})^\perp} W)^{-1} - \Pi_{\text{span}(f_{\max})}$ .*

**Lemma 4.4.**  *$\mathfrak{P}_{SEM}$  satisfies Assumption 3 with  $L(\rho) = (I_n - \rho W)^{-1}$  and  $\mathbf{z}$  a random vector distributed like  $\boldsymbol{\varepsilon}$ . Furthermore, if the distribution of  $\boldsymbol{\varepsilon}$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ , or, more generally, if  $\Pr(\boldsymbol{\varepsilon} = 0) = 0$  and the distribution of  $\boldsymbol{\varepsilon}/\|\boldsymbol{\varepsilon}\|$  is absolutely continuous w.r.t. the uniform distribution  $v_{S^{n-1}}$  on the unit sphere  $S^{n-1}$ , then  $\mathfrak{P}_{SEM}$  satisfies Assumption 2.*

<sup>30</sup>It appears that it is implicitly assumed in Martellosio (2010) that  $\boldsymbol{\varepsilon}$  is a random vector whose distribution is independent of  $\beta$ ,  $\sigma$ , and  $\rho$ , cf. Martellosio (2010), p. 155. As discussed in Remark 2.1(ii), it is also implicitly assumed in Martellosio (2010) that the distribution of  $\sigma^{-1} \Sigma_{SEM}^{-1/2}(\rho) \mathbf{u}$  is independent of  $\beta$ ,  $\sigma$ , and  $\rho$ . Note that the latter random vector is connected to  $\boldsymbol{\varepsilon}$  via multiplication by an orthogonal matrix  $U(\rho)$ , say. If  $W$  is symmetric,  $U(\rho) \equiv I_n$  holds and hence both implicit assumptions are equivalent. However, for nonsymmetric  $W$ , these two implicit assumptions will typically be compatible only if the distribution of  $\boldsymbol{\varepsilon}$  is spherically symmetric.

<sup>31</sup>Lemmata 4.2 and 4.3 actually hold without the additional assumption on the distribution of  $\boldsymbol{\varepsilon}$  made above.

Given the preceding two lemmata the main results of Section 2.2, i.e., Theorems 2.7, 2.16, and 2.18, can be immediately applied to obtain results for the spatial error model. Rather than spelling out these general results, we provide the following two corollaries for the purpose of illustration and thus do not strive for the weakest conditions. These corollaries provide, in particular, correct versions of the claims in Corollary 1 in Martellosio (2010). Recall that by the assumed  $G_X$ -invariance the rejection probabilities  $E_{\beta,\sigma,\rho}(\varphi)$  in the subsequent results do in fact neither depend on  $\beta$  nor  $\sigma$ , cf. Remark 2.3.

**Corollary 4.5.** *Given the maintained assumptions for the SEM suppose furthermore that either (i) the distribution of  $\varepsilon$  possesses a  $\mu_{\mathbb{R}^n}$ -density  $p$  that is continuous  $\mu_{\mathbb{R}^n}$ -almost everywhere and that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set, or (ii) the distribution of  $\varepsilon$  is spherically symmetric with no atom at the origin. Then for every  $G_X$ -invariant test  $\varphi$  the following statements hold:*

1. *If  $\varphi$  is continuous at  $f_{\max}$  then for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , we have  $E_{\beta,\sigma,\rho}(\varphi) \rightarrow \varphi(f_{\max})$  for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ .*
2. *Suppose  $\varphi$  satisfies  $\varphi(y) = \varphi(y + f_{\max})$  for every  $y \in \mathbb{R}^k$  (which is certainly the case if  $f_{\max} \in \text{span}(X)$ ). Then for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , we have  $E_{\beta,\sigma,\rho}(\varphi) \rightarrow E\varphi(\Lambda\varepsilon)$  for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ . The limit  $E\varphi(\Lambda\varepsilon)$  is strictly between 0 and 1 provided neither  $\varphi = 0$   $\mu_{\mathbb{R}^n}$ -almost everywhere nor  $\varphi = 1$   $\mu_{\mathbb{R}^n}$ -almost everywhere holds. [The matrix  $\Lambda$  is defined in Lemma 4.3.]*
3. *If  $\varphi$  is the indicator function of a critical region  $\Phi$ , we have for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ :*
  - *$f_{\max} \in \text{int}(\Phi)$  implies  $P_{\beta,\sigma,\rho}(\Phi) \rightarrow 1$ .*
  - *$f_{\max} \notin \text{cl}(\Phi)$  implies  $P_{\beta,\sigma,\rho}(\Phi) \rightarrow 0$ .*
  - *$f_{\max} \in \text{span}(X)$  implies  $P_{\beta,\sigma,\rho}(\Phi) \rightarrow \Pr(\Lambda\varepsilon \in \Phi)$ . The limiting probability is strictly between 0 and 1 provided neither  $\Phi$  nor its complement are  $\mu_{\mathbb{R}^n}$ -null sets.*
4. *If  $\varphi$  is the indicator function of the critical region  $\Phi_{B,\kappa}$  given by (7) with  $B$  satisfying  $\lambda_1(B) < \lambda_{n-k}(B)$  and with  $\kappa \in [\lambda_1(B), \lambda_{n-k}(B))$ , then we have for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ :*
  - *$T_B(f_{\max}) > \kappa$  implies  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow 1$ .<sup>32</sup>*
  - *$T_B(f_{\max}) < \kappa$  and  $f_{\max} \notin \text{span}(X)$  implies  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow 0$ .*
  - *$f_{\max} \in \text{span}(X)$  implies  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow \Pr(\Lambda\varepsilon \in \Phi_{B,\kappa})$ . The limiting probability is strictly between 0 and 1 provided  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$ , while it is 1 for  $\kappa = \lambda_1(B)$ .*

**Remark 4.6.** If  $\varphi = 0$  ( $= 1$ )  $\mu_{\mathbb{R}^n}$ -almost everywhere in Part 2 or in the last claim of Part 3 of the preceding corollary, then  $E_{\beta,\sigma,\rho}(\varphi) = 0$  (or  $= 1$ ) holds for all  $\beta$ ,  $\sigma$ , and  $\rho$ , and hence the same holds a fortiori for the accumulation points, see Remark 2.17(iv).

---

<sup>32</sup>Note that  $T_B(f_{\max}) > \kappa$  entails  $f_{\max} \notin \text{span}(X)$  in view of (8) and  $\kappa \geq \lambda_1(B)$ .

Parts 3 and 4 of the preceding corollary are silent on the case  $f_{\max} \in \text{bd}(\Phi) \setminus \text{span}(X)$  (recall that  $\text{span}(X) \subseteq \text{bd}(\Phi)$  holds provided  $\emptyset \neq \Phi \neq \mathbb{R}^n$ ). The next corollary provides such a result for the important critical regions  $\Phi_{B,\kappa}$  under an elliptical symmetry assumption on  $\mathfrak{P}_{SEM}$  and under the assumption of a symmetric weights matrix  $W$ . More general results without the symmetry assumption on  $W$ , without the elliptical symmetry assumption, and for more general classes of tests can of course be obtained from Theorem 2.18.

**Corollary 4.7.** *Given the maintained assumptions for the SEM suppose furthermore that the distribution of  $\varepsilon$  is spherically symmetric with no atom at the origin and that  $W$  is symmetric. Let the critical region  $\Phi_{B,\kappa}$  be given by (7). Assume  $f_{\max} \in \text{bd}(\Phi_{B,\kappa}) \setminus \text{span}(X)$  (i.e.,  $f_{\max} \notin \text{span}(X)$ ) and  $\kappa = T_B(f_{\max}) \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$  hold).*

1. *Suppose  $C_X f_{\max}$  is an eigenvector of  $B$  with eigenvalue  $\lambda$ . Then  $\lambda = T_B(f_{\max}) = \kappa$  and*

$$P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow \Pr(\mathbf{G}'\Lambda'(C_X'BC_X - \lambda C_X'C_X)\Lambda\mathbf{G} > 0) \quad (27)$$

*for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ , and for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . The limit in (27) is strictly between 0 and 1 if  $\lambda > \lambda_1(B)$ , whereas it equals 1 in case  $\lambda = \lambda_1(B)$ .*

2. *Suppose  $C_X f_{\max}$  is not an eigenvector of  $B$ . Then  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow 1/2$  for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ , and for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ .*

**Remark 4.8.** *(Some comments on Krämer (2005))* (i) Krämer (2005) considers "test statistics" of the form  $\mathbf{u}'Q_1\mathbf{u}/\mathbf{u}'Q_2\mathbf{u}$  for general matrices  $Q_1$  and  $Q_2$ . However, this ratio will then in general not be observable and thus will not be a test statistic. Fortunately, the problem disappears in the leading cases where  $Q_1$  and  $Q_2$  are such that  $\mathbf{u}'Q_i\mathbf{u} = \mathbf{y}'Q_i\mathbf{y}$ . The same problem also appears in Krämer and Zeisel (1990) and Small (1993).

(ii) The proof of the last claim in Theorem 1 of Krämer (2005) is in error, as – contrary to the claim in Krämer (2005) – the quantity  $d_1$  need not be strictly positive. This has already been noted by Martellosio (2012), Footnote 5.

(iii) Theorem 2 in Krämer (2005) is not a theorem in the mathematical sense, as it is not made precise what it means that the limiting power "is in general strictly between 0 and 1".

As discussed earlier, point-optimal invariant and locally best invariant tests are in general not immune to the zero-power trap. The next result, which is a correct version of Proposition 1 in Martellosio (2010), now provides a necessary and sufficient condition for the Cliff-Ord test (i.e.,  $B = W + W'$ ) and a point-optimal invariant test (i.e.,  $B = -\Sigma_{SEM}^{-1}(\bar{\rho})$ ) in a pure SAR-model (i.e.,  $k = 0$ ) to have limiting power equal to 1 for every choice of the critical value  $\kappa$  (excluding trivial cases). For a discussion of the problems with Proposition 1 in Martellosio (2010) see Appendix B.2. In the subsequent proposition we always have  $\lambda_1(B) < \lambda_n(B)$  as a consequence of the assumptions. We also note that the condition  $\kappa \in (\lambda_1(B), \lambda_n(B))$  in this proposition precisely corresponds to the condition that the test has size strictly between zero and one, cf. Remark 2.27. Furthermore, observe that while the statement that the limiting power (as  $\rho \rightarrow \lambda_{\max}^{-1}$ ) equals 1 for every  $\kappa \in (\lambda_1(B), \lambda_n(B))$  is in general clearly stronger than the statement that  $\alpha^*(T_B) = 0$ , Proposition 2.26 shows that these statements are in fact equivalent in the context of the following result. Finally, recall that in view of invariance and the maintained assumptions of this section the rejection probabilities do neither depend on  $\beta$  nor  $\sigma$ .

**Proposition 4.9.** *Given the maintained assumptions for the SEM, suppose that the distribution of  $\varepsilon$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$  with a density that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set. Furthermore, assume that  $k = 0$ . Let (i)  $B = W + W'$  or (ii)  $B = -\Sigma_{SEM}^{-1}(\bar{\rho})$  for some  $0 < \bar{\rho} < \lambda_{\max}^{-1}$ . Consider the rejection region  $\Phi_{B,\kappa}$  given by (7) with  $C_0 = I_n$ . Then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have in both cases (i) and (ii):  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow 1$  for every  $\kappa \in (\lambda_1(B), \lambda_n(B))$  as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ , if and only if  $f_{\max} \in \text{Eig}(B, \lambda_n(B))$ . In particular, if  $W$  is (elementwise) nonnegative and irreducible, then, for both choices of  $B$ ,  $f_{\max} \in \text{Eig}(B, \lambda_n(B))$  is equivalent to  $f_{\max}$  being an eigenvector of  $W'$ .*

The next proposition is a correct version of Lemma E.4 in Martellosio (2010); see Appendix B.2 for a discussion of the shortcomings of that lemma. It provides conditions under which the Cliff-Ord test and point-optimal invariant tests in a SEM with exogenous variables are not subject to the zero-power trap and even have limiting power equal to 1.

**Proposition 4.10.** *Given the maintained assumptions for the SEM, suppose that the distribution of  $\varepsilon$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$  with a density that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set. Suppose further that  $f_{\max} \notin \text{span}(X)$ , that  $\text{Eig}(C_X \Sigma_{SEM}(\rho) C_X', \lambda_{n-k}(C_X \Sigma_{SEM}(\rho) C_X'))$  is independent of  $0 < \rho < \lambda_{\max}^{-1}$ , and that  $n - k > 1$ . Let (i)  $B = C_X(W + W')C_X'$  and suppose that  $\lambda_1(B) < \lambda_{n-k}(B)$ , or (ii)  $B = -(C_X \Sigma_{SEM}(\bar{\rho}) C_X')^{-1}$  for some  $0 < \bar{\rho} < \lambda_{\max}^{-1}$ . Consider the rejection region  $\Phi_{B,\kappa}$  given by (7). Then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have in both cases (i) and (ii):  $P_{\beta,\sigma,\rho}(\Phi_{B,\kappa}) \rightarrow 1$  for every  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$  as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ .*

**Remark 4.11.** (i) The condition that  $\text{Eig}(C_X \Sigma_{SEM}(\rho) C_X', \lambda_{n-k}(C_X \Sigma_{SEM}(\rho) C_X'))$  is independent of  $\rho$  is easily seen to be satisfied, e.g., if  $W$  is symmetric and if  $f_{\max} \in \text{span}(X)^\perp$  (and thus, in particular, if  $k = 0$ ).

(ii) If in the preceding proposition  $W$  is symmetric and  $f_{\max} \in \text{span}(X)^\perp$  holds, then the condition  $\lambda_1(B) < \lambda_{n-k}(B)$  in case  $B = C_X(W + W')C_X' = 2C_X W C_X'$  is automatically satisfied. This can be seen as follows: Since  $f_{\max} \in \text{span}(X)^\perp$  we can represent  $f_{\max}$  as  $C_X' \gamma$  for some  $\gamma \in \mathbb{R}^{n-k}$  with  $\gamma' \gamma = 1$ . On the one hand, the largest eigenvalue of  $2C_X W C_X'$ , as the maximum of  $2\delta' C_X W C_X' \delta$  over all normalized vectors  $\delta \in \mathbb{R}^{n-k}$ , is therefore not less than  $2\gamma' C_X W C_X' \gamma = 2f_{\max}' W f_{\max} = 2\lambda_{\max}$ . On the other hand, noting that  $\|C_X' \delta\| = \|\delta\|$ , the maximum of  $2\delta' C_X W C_X' \delta$  over all normalized vectors  $\delta \in \mathbb{R}^{n-k}$  is not larger than the maximum of  $v' W v$  over all normalized vectors  $v \in \mathbb{R}^n$ , which shows that the largest eigenvalue of  $2C_X W C_X'$  is equal to  $2\lambda_{\max}$ . Because  $\lambda_{\max}$  as the largest eigenvalue of  $W$  has algebraic multiplicity 1 by the assumptions in this section and since  $2C_X W C_X'$  is symmetric, we see that the algebraic multiplicity of  $2\lambda_{\max}$  as an eigenvalue of  $2C_X W C_X'$  must also be 1. But then  $\lambda_1(B) < \lambda_{n-k}(B)$  follows since  $n - k > 1$  has been assumed in the proposition.

(iii) If  $n - k = 1$  or if  $n - k > 1$ , but  $\lambda_1(B) = \lambda_{n-k}(B)$  holds for  $B = C_X(W + W')C_X'$ , then the test statistic degenerates to a constant (and the proposition trivially holds as  $(\lambda_1(B), \lambda_{n-k}(B))$  is then empty).

## 4.2 Spatial lag models

Let  $X$  be as in Section 2.1, let  $W$  be as in Section 4.1, and consider the spatial lag model (SLM) of the form

$$\mathbf{y} = \rho W \mathbf{y} + X \beta + \sigma \varepsilon, \quad (28)$$



where  $\beta \in \mathbb{R}^k$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ , and  $0 < \sigma < \infty$ , and where  $\varepsilon$  is a mean zero random vector with covariance matrix  $I_n$ . As in Section 4.1, we assume that the distribution of  $\varepsilon$  is a fixed distribution independent of  $\beta \in \mathbb{R}^k$ ,  $\sigma \in (0, \infty)$ , and  $\rho \in [0, \lambda_{\max}^{-1})$ . *The above are the maintained assumptions for the SLM considered in this section.* Because the SLM and the SEM have the same covariance structure, a simple consequence of Lemma 4.2 is that also the parameters of the SLM are identifiable. For  $\rho \in [0, \lambda_{\max}^{-1})$  we can rewrite the above equation as

$$\mathbf{y} = (I_n - \rho W)^{-1}(X\beta + \sigma\varepsilon). \quad (29)$$

Obviously, in case  $k = 0$  the spatial lag model of order one coincides with the SAR(1) model. For  $k > 0$ , however, the SLM does *not* fit into the general framework of Section 2.2 of the present paper. In particular, while the problem of testing  $\rho = 0$  versus  $\rho > 0$  is still invariant under the group  $G_0$ , it is typically no longer invariant under the larger group  $G_X$ . Nevertheless we can establish the following result which is similar in spirit to Theorem 2.7. In the following result,  $P_{\beta, \sigma, \rho}$  denotes the distribution of  $\mathbf{y}$  given by (29) under the parameters  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and  $\rho \in [0, \lambda_{\max}^{-1})$  and  $E_{\beta, \sigma, \rho}$  denotes the corresponding expectation operator.

**Theorem 4.12.** *Given the maintained assumptions for the SLM, assume furthermore that the distribution of  $\varepsilon$  does not put positive mass on a proper affine subspace of  $\mathbb{R}^n$ . Let  $\varphi$  be a  $G_0$ -invariant test. If  $\varphi$  is continuous at  $f_{\max}$  then for every  $\beta \in \mathbb{R}^k$  and  $0 < \sigma < \infty$  we have  $E_{\beta, \sigma, \rho}(\varphi) \rightarrow \varphi(f_{\max})$  for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ . In particular, if  $\varphi$  is the indicator function of a critical region  $\Phi$  we have for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , and as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ :*

- $f_{\max} \in \text{int}(\Phi)$  implies  $P_{\beta, \sigma, \rho}(\Phi) \rightarrow 1$ .
- $f_{\max} \notin \text{cl}(\Phi)$  implies  $P_{\beta, \sigma, \rho}(\Phi) \rightarrow 0$ .

The above result provides a correct version of the first and third claim in Proposition 2 in Martellosio (2010), the proofs of which in Martellosio (2010) suffer from the same problems as the proofs of the corresponding parts of MT1. The second claim in Proposition 2 in Martellosio (2010) is incorrect for the same reasons as is the second part of MT1. While Theorems 2.16 and 2.18 provide correct versions of the second claim of MT1, these results can not directly be used in the context of the SLM as this model does not fit into the framework of Section 2.2 as noted above. We do not investigate this issue any further here.<sup>33</sup>

### 4.3 Indistinguishability by invariant tests in spatial regression models

We close our discussion of spatial regression models by applying the results on indistinguishability developed in Section 2.3 to these models. It turns out that a number of results in Martellosio (2010) (namely all parts of Proposition 3, 4, and 5 that are based on degeneracy of the test statistic) as well as the first part of the theorem in Martellosio (2011b) are consequences of an identification problem in the distribution of the maximal invariant statistic (more precisely, an identification problem in the "reduced" experiment). Theorem 2.30 and Corollary 2.31 thus provide a simple and systematic way to recognize when this identification problem occurs.

---

<sup>33</sup>Under additional restrictive assumptions (such as  $\text{span}((I_n - \rho W)^{-1}X) \subseteq \text{span}(X)$  for every  $\rho \in [0, \lambda_{\max}^{-1})$ ) invariance w.r.t.  $G_X$  can again become an appropriate assumption on a test statistic and a version of Theorem 2.16 can then be produced. We abstain from pursuing this any further.

Consider the SEM with the maintained assumptions of Section 4.1 and additionally assume *for this paragraph only* that the distribution of the error  $\varepsilon$  is spherically symmetric. As shown in Section 2.3, the condition for the identification problem in the reduced experiment to occur, entailing a constant power function for any  $G_X$ -invariant (even for any  $G_X^+$ -invariant) test, is then that  $C_X \Sigma_{SEM}(\rho) C_X'$  is a multiple of  $I_{n-k}$  for every  $\rho \in (0, \lambda_{\max}^{-1})$ . As can be seen from Lemma C.5 in Appendix C, a sufficient condition for this is that  $\text{span}(X)^\perp$  is contained in an eigenspace of  $\Sigma_{SEM}(\rho)$  for every  $\rho \in (0, \lambda_{\max}^{-1})$ , a condition that appears in Proposition 3 of Martellosio (2010), which is a statement about point-optimal invariant and locally best invariant tests. Thus the corresponding part of this proposition is an immediate consequence of Corollary 2.31; moreover, and in contrast to this proposition in Martellosio (2010), it now follows that this result holds more generally for *any*  $G_X$ -invariant (even any  $G_X^+$ -invariant) test and that the Gaussianity assumption in this proposition can be weakened to elliptical symmetry. In a similar way, Propositions 4 and 5 in Martellosio (2010) make use of the conditions that  $W$  is symmetric and  $\text{span}(X)^\perp$  is contained in an eigenspace of  $W$ . In the subsequent lemma we show that the condition that  $\text{span}(X)^\perp$  is contained in an eigenspace of  $W'$  is already sufficient for  $C_X \Sigma_{SEM}(\rho) C_X'$  to be a multiple of  $I_{n-k}$  for every  $\rho \in (0, \lambda_{\max}^{-1})$ . Thus the subsequent lemma combined with Corollary 2.31 establishes, in particular, the respective parts of Propositions 4 and 5 in Martellosio (2010). The preceding comments are of some importance as there are several problems with Propositions 3, 4, and 5 in Martellosio (2010) which are discussed in Appendix B.3.

**Lemma 4.13.** *Let  $W$  be a weights matrix as in Section 4.1 and let  $X$  be an  $n \times k$  matrix ( $n > k$ ) such that*

$$\text{span}(X)^\perp \subseteq \text{Eig}(W', \lambda) \quad (30)$$

*is satisfied for some eigenvalue  $\lambda \in \mathbb{R}$  of  $W'$ . Then*

$$\Pi_{\text{span}(X)^\perp} (I_n - \rho W)^{-1} = (1 - \rho \lambda)^{-1} \Pi_{\text{span}(X)^\perp}$$

*and*

$$C_X \Sigma_{SEM}(\rho) C_X' = (1 - \rho \lambda)^{-2} I_{n-k}$$

*hold for every  $0 \leq \rho < \lambda_{\max}^{-1}$ .*

In the following example we show that the first half of the theorem in Martellosio (2011b) is a special case of Remark 2.33 following Corollary 2.31 combined with the preceding lemma.

**Example 4.1.** (i) Consider the SEM with the maintained assumptions of Section 4.1. Suppose that  $W$  is an  $n \times n$  ( $n \geq 2$ ) equal weights matrix, i.e.,  $w_{ij}$  is constant for  $i \neq j$  and zero else and that  $\text{span}(X)$  contains the intercept. Without loss of generality we assume  $w_{ij} = 1$  for  $i \neq j$ . Clearly,  $W$  is symmetric and has the eigenvalues  $\lambda_1(W) = \dots = \lambda_{n-1}(W) = -1$  and  $\lambda_{\max} = \lambda_n(W) = n - 1$ . The eigenspace corresponding to  $\lambda_{\max}$  is spanned by the eigenvector  $f_{\max} = n^{-1/2}(1, \dots, 1)'$  and the other eigenspace consists of all vectors orthogonal to  $f_{\max}$ . Since every element of  $\text{span}(X)^\perp$  is orthogonal to  $(1, \dots, 1)'$  we have

$$\text{span}(X)^\perp \subseteq \text{Eig}(W, -1) = \text{Eig}(W', -1).$$

Therefore, by Lemma 4.13 together with Remark 2.33, the power function of every  $G_X^+$ -invariant test must be constant.

(ii) Consider next the SLM with the maintained assumptions of Section 4.2 with the same weights matrix and the same design matrix as in (i). Observe that  $W$  can be written as  $W =$

$nf_{\max}f'_{\max} - I_n$ , a matrix which obviously maps  $\text{span}(X)$  into  $\text{span}(X)$  as the intercept has been assumed to be an element of  $\text{span}(X)$ . Consequently, also  $I_n - \rho W$  maps  $\text{span}(X)$  into  $\text{span}(X)$  for every  $\rho \in [0, \lambda_{\max}^{-1})$ . Because  $I_n - \rho W$  is nonsingular for  $\rho$  in that range, it follows that this mapping is onto and furthermore that also  $(I_n - \rho W)^{-1}$  maps  $\text{span}(X)$  into  $\text{span}(X)$  in a bijective way. As a consequence, the mean of  $\mathbf{y}$ , which equals  $(I_n - \rho W)^{-1}X\beta$ , is an element  $X\gamma(\beta, \rho)$ , say, of  $\text{span}(X)$  for every  $\rho \in [0, \lambda_{\max}^{-1})$  and  $\beta \in \mathbb{R}^k$ .<sup>34</sup> Let  $\varphi$  be any  $G_X^+$ -invariant test. Then by  $G_X^+$ -invariance we have

$$E\varphi(\mathbf{y}) = E\varphi(X\gamma(\beta, \rho) + \sigma(I_n - \rho W)^{-1}\boldsymbol{\varepsilon}) = E\varphi(\sigma(I_n - \rho W)^{-1}\boldsymbol{\varepsilon}) = E\varphi(X\beta + \sigma(I_n - \rho W)^{-1}\boldsymbol{\varepsilon}),$$

which coincides with the power function in a SEM as in (i) above and thus is independent of  $\beta$ ,  $\sigma$ , and  $\rho$ , showing that the power function of any  $G_X^+$ -invariant test in the SLM considered here is constant.  $\square$

## 5 An application to time-series regression models

In this section we briefly comment on the case where the error vector  $\mathbf{u}$  in (1) has covariance matrix  $\Sigma(\rho)$  for  $\rho \in [0, 1)$  with the  $(i, j)$ -th element of  $\Sigma(\rho)$  given by  $\rho^{|i-j|}$  (Case I) or  $(-\rho)^{|i-j|}$  (Case II). Clearly, Case I corresponds to testing against positive autocorrelation, while Case II corresponds to testing against negative autocorrelation. More precisely, in both cases we assume that  $\mathbf{u}$  is distributed as  $\sigma\Sigma^{1/2}(\rho)\boldsymbol{\varepsilon}$ , where  $\boldsymbol{\varepsilon}$  has mean zero, has covariance matrix  $I_n$ , and has a fixed distribution that is spherically symmetric (and hence does not depend on any parameters); in particular, Assumption 3 is maintained. Furthermore, assume that  $\Pr(\boldsymbol{\varepsilon} = 0) = 0$ . *We shall refer to these assumptions as the maintained assumptions of this section.* This framework clearly covers the case where the vector  $\mathbf{u}$  is a segment of a Gaussian stationary autoregressive process of order 1. In Case I it is now readily verified that Assumption 1 holds with  $e = n^{-1/2}(1, \dots, 1)'$ , while in Case II this assumption is satisfied with  $e = n^{-1/2}(-1, 1, \dots, (-1)^n)'$ . The validity of Assumption 2 then follows from Proposition 2.6. Furthermore, Assumption 4 (more precisely, the equivalent condition given in Lemma 2.14) has been shown to be satisfied in Case I as well as in Case II in Lemma G.1 of Preinerstorfer and Pötscher (2013), where the form of the matrix  $V$  (denoted by  $D$  in that reference) is also given; this lemma also establishes condition (18) in view of the fact that obviously  $\lambda_n(\Sigma(\rho)) \rightarrow n$  for  $\rho \rightarrow 1$  (in Case I as well as in Case II). We thus immediately get the following result as a special case of the results in Section 2.2:

**Corollary 5.1.** *Suppose the maintained assumptions hold. Let  $e$  denote  $n^{-1/2}(1, \dots, 1)'$  in Case I while it denotes  $n^{-1/2}(-1, 1, \dots, (-1)^n)'$  in Case II.*

1. *Then every  $G_X$ -invariant test  $\varphi$  satisfies the conclusions 1.-4. of Corollary 4.5 subject to replacing  $\lambda_{\max}^{-1}$  by 1,  $f_{\max}$  by  $e$ , and where  $\Lambda$  now represents a square-root of the matrix  $D$  given in Lemma G.1 of Preinerstorfer and Pötscher (2013).*
2. *Let the critical region  $\Phi_{B, \kappa}$  be given by (7). Assume  $e \in \text{bd}(\Phi_{B, \kappa}) \setminus \text{span}(X)$  (i.e.,  $e \notin \text{span}(X)$  and  $\kappa = T_B(e) \in [\lambda_1(B), \lambda_{n-k}(B))$  with  $\lambda_1(B) < \lambda_{n-k}(B)$  hold). Then:*

---

<sup>34</sup>Compare Footnote 2 in Martellosio (2011b), where the author attempts to justify invariance in case of a spatial lag model. The argument given there to show that  $(I - \rho W)^{-1}$  for  $W$  an equal weights matrix maps  $\text{span}(X)$  into itself, however, does not make sense as it is based on an incorrect expression for  $E(\mathbf{y})$ , which is incorrectly given as  $\Delta_\rho X$ .

(i) Suppose  $C_X e$  is an eigenvector of  $B$  with eigenvalue  $\lambda$ . Then  $\lambda = T_B(e) = \kappa$  and

$$P_{\beta, \sigma, \rho}(\Phi_{B, \kappa}) \rightarrow \Pr(\mathbf{G}' \Lambda' (C_X' B C_X - \lambda C_X' C_X) \Lambda \mathbf{G} > 0) \quad (31)$$

for  $\rho \rightarrow 1$ ,  $\rho \in [0, 1)$ , and for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ , where  $\mathbf{G}$  is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . The limit in (31) is strictly between 0 and 1 if  $\lambda > \lambda_1(B)$ , whereas it equals 1 in case  $\lambda = \lambda_1(B)$ .

(ii) Suppose  $C_X e$  is not an eigenvector of  $B$ . Then  $P_{\beta, \sigma, \rho}(\Phi_{B, \kappa}) \rightarrow 1/2$  for  $\rho \rightarrow 1$ ,  $\rho \in [0, 1)$ , and for every  $\beta \in \mathbb{R}^k$ ,  $0 < \sigma < \infty$ .

The proof of the corollary is similar to the proof of Corollaries 4.5 and 4.7 and consists of a straightforward application of Theorems 2.7, 2.16, and Corollary 2.23, noting that condition (18) has been verified in Lemma G.1 of Preinerstorfer and Pötscher (2013). At the expense of arriving at a more complicated result, some of the maintained assumptions like the spherical symmetry assumption could be weakened, while nevertheless allowing the application of the results in Section 2.2. In the literature often the alternative parameterization  $\zeta^2 (1 - \rho^2)^{-1} \Sigma(\rho)$  for the covariance matrix of  $\mathbf{u}$  is used, which just amounts to parametrizing  $\sigma^2$  as  $\zeta^2 (1 - \rho^2)^{-1}$ . In view of Remark 2.3 and  $G_X$ -invariance of the tests considered, such an alternative reparameterization has no effect on the results in this section at all.

Even after specializing to the Gaussian case, the preceding corollary provides a substantial generalization of a number of results in the literature in that (i) it allows for general  $G_X$ -invariant tests rather than discussing some specific tests, and (ii) provides explicit expressions for the limiting power also in the case where the limit is neither zero nor one: Krämer (1985) appears to have been the first to notice that the zero-power trap can arise for the Durbin-Watson test in that he showed that the limiting power (as the autocorrelation tends to 1) of the Durbin-Watson test can be zero when one considers a linear regression model without an intercept and with the errors following a Gaussian autoregressive process of order one. More precisely, he established that in this model the limiting power is zero (is one) if – in our notation – the vector  $e$  is outside the closure (is inside the interior) of the rejection region of the Durbin-Watson test.<sup>35</sup> Based on numerical results, he also noted that the zero-power trap does not seem to arise in models that contain an intercept. Subsequently, Zeisel (1989) showed that indeed in models with an intercept the limiting power of the Durbin-Watson test (except in degenerate cases) is always strictly between zero and one.<sup>36</sup> The results in Krämer (1985) and Zeisel (1989) just mentioned are extended in Krämer and Zeisel (1990) from the Durbin-Watson test to tests that can be expressed as ratios of quadratic forms, see also Small (1993).<sup>37</sup> [We note that Krämer (1985) and Krämer and Zeisel (1990) additionally also consider the case where the autocorrelation tends to  $-1$ .] All these results can be easily read off from Part 1 of our Corollary 5.1. The analysis in Krämer (1985), Zeisel (1989), and Krämer and Zeisel (1990) always excludes a particular case, which is treated in Löbus and Ritter (2000) for the Durbin-Watson test. This result is again easily seen to be a special case of Part 2 of our Corollary 5.1. Furthermore, Zeisel (1989) shows that for any sample size  $n$  and number of regressors  $k < n$  a design matrix exists such that zero-power trap arises. For a systematic investigation of the set of regressors for which the zero-power trap occurs see Preinerstorfer (2014).

<sup>35</sup>We note that the words "inside" and "outside" in the Corollary of Krämer (1985) should be interchanged.

<sup>36</sup>The argument in Zeisel (1989) tacitly makes use of the Portmanteau theorem in deriving the formula for the limiting rejection probability without providing the necessary justification. For a more complete proof along the same lines as the one in Zeisel (1989) see Löbus and Ritter (2000).

<sup>37</sup>See Remark 4.8(i).

## A Comments on and counterexamples to Theorem 1 in Martellosio (2010)

As already mentioned in Section 2.2, the first and third claim in MT1 are correct, but the proof of these statements as given in Martellosio (2010) is not (cf. also Mynbaev (2012)). To explain the mistake, we assume for simplicity that  $\mathbf{u}$  is Gaussian. With this additional assumption the model satisfies all the requirements imposed in Martellosio (2010), page 154 (cf. Remark 2.1 above). The proof of MT1 in Martellosio (2010) is given for  $\beta$  arbitrary and  $\sigma = 1$ . Set  $\beta = 0$  for simplicity. In the proof of MT1 it is argued that the density of  $\mathbf{y}$  tends, as  $\rho \rightarrow a$ , to a degenerate "density" which is supported on a set that simplifies to the eigenspace of  $\Sigma^{-1}(a-)$  corresponding to its smallest eigenvalue in the case  $\beta = 0$  considered here. However, for  $\rho \in [0, a)$ , the density of  $\mathbf{y}$  is

$$f(y) = (2\pi)^{-n/2} (\det(\Sigma^{-1}(\rho)))^{1/2} \exp \left\{ -\frac{1}{2} y' \Sigma^{-1}(\rho) y \right\}.$$

As  $\rho \rightarrow a$  we have  $\det(\Sigma^{-1}(\rho)) \rightarrow 0$  in view of the assumption  $\text{rank}(\Sigma^{-1}(a-)) = n - 1$ . Furthermore,  $\exp \left\{ -\frac{1}{2} y' \Sigma^{-1}(\rho) y \right\} \rightarrow \exp \left\{ -\frac{1}{2} y' \Sigma^{-1}(a-) y \right\}$  (even uniformly on compact subsets). Therefore the density converges to zero everywhere (and even uniformly on compact subsets). In particular, it does not tend to a degenerate "density" supported on the eigenspace of  $\Sigma^{-1}(a-)$  corresponding to its smallest eigenvalue in any suitable way. Note that  $P_{0,1,\rho}$  does also not converge weakly as  $\rho \rightarrow a$  as the sequence  $P_{0,1,\rho_m}$  for any  $\rho_m \rightarrow a$  is obviously not tight. This shows that the proof in Martellosio (2010) is incorrect. Furthermore, the concentration effect discussed after the theorem in Martellosio (2010) simply does not occur in the way as claimed. In fact, the direct opposite happens: the distributions stretch out, i.e., all of the mass "escapes to infinity".

We next turn to the second claim in MT1 and show by two simple counterexamples that this claim is not correct.<sup>38</sup> The first example below is based on the following simple observation: Suppose  $k = 0$ , the testing problem satisfies all assumptions of MT1, and we can find an invariant rejection region  $\tilde{\Phi}$  of size  $\alpha$ ,  $0 < \alpha < 1$ , with  $e \in \text{int}(\tilde{\Phi})$ . The (correct) first claim of MT1 then implies that the limiting power of  $\tilde{\Phi}$  is 1. Now define  $\Phi = \tilde{\Phi} \setminus \text{span}(e)$  and observe that  $\Phi$  is again invariant and that  $\Phi$  and  $\tilde{\Phi}$  have the same rejection probabilities as they differ only by a  $\mu_{\mathbb{R}^n}$ -null set and the family  $\mathfrak{P}$  is dominated by  $\mu_{\mathbb{R}^n}$  under the assumptions in Martellosio (2010). Now  $e \in \text{bd}(\Phi)$  holds, but the limiting power of  $\Phi$  is obviously 1. A concrete counterexample is as follows:

**Example A.1.** Assume that the elements  $P_{\beta,\sigma,\rho}$  of the family  $\mathfrak{P}$  are Gaussian, i.e.,  $\mathfrak{P}$  satisfies Assumption 3 with  $\mathbf{z}$  a standard normally distributed vector (and without loss of generality we may set  $L(\rho) = \Sigma^{1/2}(\rho)$ ). For simplicity we consider the case without regressors (i.e., we assume  $k = 0$  and thus  $\beta = 0$  holds by our conventions). Let

$$\Sigma(\rho) = I_n + (1 - \rho)^{-1} \rho e e'$$

for every  $\rho \in [0, 1)$  where  $e$  is normalized. Clearly,  $\Sigma(\rho)$  is symmetric and positive definite for  $\rho \in [0, 1)$  and  $\Sigma(0) = I_n$  holds. Observe that  $\Sigma^{-1}(\rho) = I_n - \rho e e'$ , and thus  $\Sigma^{-1}(1-) = I_n - e e'$ , which has rank  $n - 1$ . The family  $\mathfrak{P}$  hence clearly satisfies all the assumptions for MT1 imposed in Martellosio (2010), cf. Remark 2.1 in Section 2.1 above. Now fix an arbitrary  $\alpha \in (0, 1)$  and choose

---

<sup>38</sup>Mynbaev (2012) also claims to provide counterexamples to the second claim in MT1. However, strictly speaking, these examples are not counterexamples as the tests constructed always have either size 0 or 1, a case ruled out in the main body of Martellosio (2010).

a rejection region  $\tilde{\Phi} \in \mathcal{B}(\mathbb{R}^n)$  that is (i) invariant w.r.t.  $G_0$ , (ii) satisfies  $P_{0,1,0}(\tilde{\Phi}) = \alpha$  (and thus  $P_{0,\sigma,0}(\tilde{\Phi}) = \alpha$  for every  $0 < \sigma < \infty$  by  $G_0$ -invariance), and (iii)  $e \in \text{int}(\tilde{\Phi})$ . [For example, choose  $M$  equal to a spherical cap on the unit sphere  $S^{n-1}$  centered at  $e$  such that  $M$  has measure  $\alpha/2$  under the uniform distribution on  $S^{n-1}$ , and set  $\tilde{\Phi} = \{\gamma y : \gamma \neq 0, y \in M\}$ .] From Remark 2.8(i) we obtain that the limiting power of  $\tilde{\Phi}$  is 1. [The assumptions of Theorem 2.7 are obviously satisfied in view of Lemma 2.5 and Proposition 2.6.] We now define a new rejection region  $\Phi = \tilde{\Phi} \setminus \text{span}(e)$ . Clearly,  $\Phi$  is also  $G_0$ -invariant, and  $\Phi$  and  $\tilde{\Phi}$  have the same rejection probabilities since  $\text{span}(e)$  is an  $\mu_{\mathbb{R}^n}$ -null set (as we have assumed  $n \geq 2$ ) and the elements of  $\mathfrak{P}$  are absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ . However, now  $e \in \text{bd}(\Phi)$  holds, showing that the second claim in MT1 is incorrect. A similar example, starting with the rejection region  $\Psi = \mathbb{R}^n \setminus \tilde{\Phi}$ , where  $\tilde{\Phi}$  is as before, and then passing to  $\Psi = \mathbb{R}^n \setminus \Phi$  provides an example where  $e \in \text{bd}(\Psi)$  holds, but the limiting power is zero.  $\square$

The argument underlying this counterexample works more generally for any covariance model  $\Sigma(\cdot)$  that satisfies the assumptions of Theorem 2.7, and thus, in particular, for spatial models.

While the rejection region  $\Phi$  constructed in the preceding example certainly provides a counterexample to the second claim in MT1, one could argue that it is somewhat artificial since  $\Phi$  can be modified by a  $\mu_{\mathbb{R}^n}$ -null set into the rejection region  $\tilde{\Phi}$  which does not have  $e$  on its boundary. One could therefore ask if there is a more genuine counterexample to the second claim of MT1 in the sense that the rejection region in such a counterexample can not be modified by a  $\mu_{\mathbb{R}^n}$ -null set in such a way that the modified region does not have  $e$  on its boundary. This is indeed the case as shown by the subsequent example.

**Example A.2.** Consider the same model as in the previous example, except that we now assume  $n = 2$  and  $\Sigma(\rho)$  is given by

$$\Sigma(\rho) = I_n + (1 - \rho)^{-1} \rho e(\rho) e'(\rho)$$

for  $\rho \in [0, 1)$  where  $e(\rho) = (\cos(\phi(\rho)), \sin(\phi(\rho)))'$  with  $\phi$  a strictly monotone and continuous function on  $[0, 1)$  satisfying  $\phi(0) = 0$  and  $\phi(1-) = \pi/2$ . Again  $\Sigma(\rho)$  is symmetric and positive definite for  $\rho \in [0, 1)$  and  $\Sigma(0) = I_n$  holds. Observe that  $\Sigma^{-1}(\rho) = I_n - \rho e(\rho) e'(\rho)$  holds and thus  $\Sigma^{-1}(1-) = I_n - e e'$  where  $e = (0, 1)'$ . Obviously,  $\Sigma^{-1}(1-)$  has rank  $n - 1$ . Again the family  $\mathfrak{P}$  satisfies all the assumptions for MT1 imposed in Martellosio (2010). Consider the rejection region  $\Phi = \{y \in \mathbb{R}^2 : y_1 y_2 \geq 0\}$  which is  $G_0$ -invariant. The rejection probability under the null is always equal to  $1/2$ . Furthermore,  $e \in \text{bd}(\Phi)$  holds (and obviously there is no modification by a  $\mu_{\mathbb{R}^n}$ -null set such that  $e \notin \text{bd}(\Phi)$ ). We next show that  $P_{0,1,\rho}(\Phi)$  converges to 1 for  $\rho \rightarrow 1$  under a suitable choice of the function  $\phi$ : By  $G_0$ -invariance,

$$P_{0,1,\rho}(\Phi) = Q_{0,(1-\rho)I_n + \rho e(\rho) e'(\rho)}(\Phi) \tag{32}$$

where  $Q_{0,\Omega}$  denotes the Gaussian measure on  $\mathbb{R}^n$  with mean zero and variance covariance matrix  $\Omega$ . Now for fixed  $\eta$ ,  $0 < \eta < 1$ , we have that  $e(\eta) \in \text{int}(\Phi)$  because of strict monotonicity of  $\phi$ . Furthermore,

$$Q_{0,(1-\rho)I_n + \rho e(\eta) e'(\eta)}(\Phi) \rightarrow Q_{0,e(\eta) e'(\eta)}(\Phi) \geq Q_{0,e(\eta) e'(\eta)}(\text{span}(e(\eta))) = 1$$

because  $Q_{0,(1-\rho)I_n + \rho e(\eta) e'(\eta)}$  converges to  $Q_{0,e(\eta) e'(\eta)}$  weakly, and because

$$Q_{0,e(\eta) e'(\eta)}(\text{bd}(\Phi)) = Q_{0,e(\eta) e'(\eta)}(\text{bd}(\Phi) \cap \text{span}(e(\eta))) = Q_{0,e(\eta) e'(\eta)}(\{0\}) = 0.$$

It is now obvious that if  $\phi(\rho)$  converges to  $\pi/2$  sufficiently slowly, we can also achieve that (32) converges to 1 as  $\rho \rightarrow 1$ . Furthermore, we also conclude that the invariant rejection region  $\Psi =$

$\mathbb{R}^2 \setminus \Phi$ , which also has rejection probability  $1/2$  under the null, provides an example where  $e \in \text{bd}(\Psi)$  holds, but the limiting power is zero.  $\square$

Similar counterexamples to the second claim in MT1 can also be constructed when regressors are present (except if  $n = k + 1$ ).<sup>39</sup>

## B Comments on further results in Martellosio (2010)

In this section we comment on problems in some results in Martellosio (2010) that have not been discussed so far. We also discuss if and how these problems can be fixed.

### B.1 Comments on Lemmata D.2 and D.3 in Martellosio (2010)

Here we discuss problems with Lemmata D.2 and D.3 in Martellosio (2010) which are phrased in a spatial error model context. Correct versions of these lemmata, which furthermore are also not restricted to spatial regression models, have been given in Section 2.2.4 above. Both lemmata in Martellosio (2010) concern the quantity  $\alpha^*$ , which is defined on p. 165 of Martellosio (2010) as follows:

”For an exact invariant test of  $\rho = 0$  against  $\rho > 0$  in a SAR(1) model,  $\alpha^*$  is the infimum of the set of values of  $\alpha \in (0, 1]$  such that the limiting power does not vanish.”

In this definition  $\alpha$  denotes a generic symbol for the size of the test. Taken literally, the definition refers to one test only and hence does not make sense (as there is then only one associated value of  $\alpha$ ). From later usage of this definition in Martellosio (2010), it seems that the author had in mind a *family* of tests (rejection regions) like  $\Phi_\kappa = \{y \in \mathbb{R}^n : T(y) > \kappa\}$ , where  $T$  is a test statistic. Interpreting Martellosio’s definition this way, it is clear that under the assumptions made in Martellosio (2010) (see Remark 2.1(ii) and Remark 2.3 above) his  $\alpha^*$  coincides with  $\alpha^*(T)$  defined in (23).

Lemma D.2 of Martellosio (2010), p. 181, then reads as follows:

”Consider a model  $G(X\beta, \sigma^2[(I - \rho W')(I - \rho W)]^{-1})$ , where  $G(\mu, \Gamma)$  denotes some multivariate distribution with mean  $\mu$  and variance matrix  $\Gamma$ . When an invariant critical region for testing  $\rho = 0$  against  $\rho > 0$  is in form (9) [i.e., is of the form  $\{y \in \mathbb{R}^n : T(y) > \kappa\}$  for some univariate test statistic  $T$ ], and is such that  $f_{\max}$  is not contained in its boundary,  $\alpha^* = \Pr(T(\mathbf{z}) > T(f_{\max}); \mathbf{z} \sim G(0, I))$ .”

The statement of this lemma as well as its proof are problematic for the following reasons:

---

<sup>39</sup>The case  $n = k + 1$  is somewhat trivial as we now explain: If  $n = k + 1$ , every  $G_X$ -invariant test  $\varphi$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere constant. [To see this observe that  $\text{span}(X)$  is a  $\mu_{\mathbb{R}^n}$ -null set and that every element of  $\text{span}(X)^\perp$  is of the form  $\lambda b$  for a fixed vector  $b$  and hence  $\varphi(y) = \varphi(\Pi_{\text{span}(X)^\perp} y) = \varphi(\lambda b) = \varphi(b)$  holds whenever  $\lambda \neq 0$ , i.e., whenever  $y \notin \text{span}(X)$ . Additionally, note that  $\varphi$  is constant on  $\text{span}(X)$ .] Consequently,  $\varphi$  has a constant power function if (i) the family of probability measures in (2) is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ , or if (ii) this family is an elliptically symmetric family (to see this in case  $\Pr(\mathbf{z} = 0) = 0$  use the argument given in Remark 2.17(iv); in case  $\Pr(\mathbf{z} = 0) > 0$  combine this argument with Remark 3.1(vi)). In particular, if  $\varphi$  is non-randomized, it is then a trivial test in that its size and power are either both zero or one, provided (i) holds or (ii) holds with  $\Pr(\mathbf{z} = 0) = 0$ .

1. The lemma makes a statement about  $\alpha^*$ , which is a quantity that depends not only on one specific critical region, but on a *family* of critical regions corresponding to a *family* of critical values  $\kappa$  against which the test statistic is compared. The critical region usually depends on  $\kappa$  and so does its boundary (cf. Proposition 2.11). Therefore, the assumption "...  $f_{\max}$  is not contained in its [the invariant critical region's] boundary..." has little meaning in this context as it is not clear to which one of the many rejection regions the statement refers to. [Alternatively, if one interprets the statement of the lemma as requiring  $f_{\max}$  not to be contained in the boundary of *every* rejection region in the family considered, this leads to a condition that typically will never be satisfied.]
2. The proof of the lemma is based on Corollary 1 in Martellosio (2010), the proof of which is incorrect as it is based on the incorrect Theorem 1 of Martellosio (2010).
3. The proof implicitly uses a continuity assumption on the cumulative distribution function of the test statistic under the null at the point  $T(f_{\max})$  which is not satisfied in general.

Next we turn to Lemma D.3 in Martellosio (2010), which reads:

"Consider a test that, in the context of a spatial error model with symmetric  $W$ , rejects  $\rho = 0$  for small values of a statistic  $\nu' B \nu$ , where  $B$  is an  $(n-k) \times (n-k)$  known symmetric matrix that does not depend on  $\alpha$ , and  $\nu$  is as defined in Section 2.2. Provided that  $f_{\max} \notin \text{bd}(\Phi)$ ,  $\alpha^* = 0$  if and only if  $C f_{\max} \in E_1(B)$ , and  $\alpha^* = 1$  if and only if  $C f_{\max} \in E_{n-k}(B)$ ."

Here  $\alpha$  refers to the size of the test,  $\nu$  is given by  $\text{sign}(y_i) C y / \|C y\|$  for some fixed  $i \in \{1, \dots, n\}$ , and  $\Phi$  is not explicitly defined, but presumably denotes a rejection region corresponding to the test statistic  $\nu' B \nu$ . [Although the test statistic is not defined whenever  $C y = 0$ , this does not pose a severe problem here since Martellosio (2010) considers only absolutely continuous distributions and since he assumes  $k < n$ ; cf Remark 2.13. Note furthermore that the factor  $\text{sign}(y_i)$  is irrelevant here.] Furthermore,  $E_1(B)$  ( $E_{n-k}(B)$ ) denotes the eigenspace corresponding to the smallest (largest) eigenvalue of  $B$ , and  $C$  in Martellosio (2010) stands for  $C_X$ . The statement of the lemma and its content are inappropriate for the following reasons:

1. The proof of this lemma is based on Lemma D.2 of Martellosio (2010) which is invalid as discussed above.
2. Again, as in the statement of Lemma D.2 of Martellosio (2010), the author assumes that "...  $f_{\max} \notin \text{bd}(\Phi)$  ...", which is not meaningful, as the boundary typically depends on the critical value.
3. The above lemma in Martellosio (2010) requires  $W$  to be symmetric (although this is actually not used in the proof). Nevertheless, it is later applied to nonsymmetric weights matrices in the proof of Proposition 1 in Martellosio (2010).

As a point of interest we note that naively applying Lemma D.3 in Martellosio (2010) to the case where  $B$  is a multiple of the identity matrix  $I_{n-k}$  leads to the contradictory statement  $0 = \alpha^* = 1$ . However, in case  $B$  is a multiple of  $I_{n-k}$ , the test statistic degenerates, and thus the size of the test is 0 or 1, a case that is ruled out in Martellosio (2010) from the very beginning.



## B.2 Comments on Proposition 1 and Lemma E.4 in Martellosio (2010)

Proposition 1 in Martellosio (2010) considers the pure SAR(1) model, i.e.,  $k = 0$  is assumed. This proposition reads as follows:

”Consider testing  $\rho = 0$  against  $\rho > 0$  in a pure SAR(1) model. The limiting power of the Cliff-Ord test [cf. eq. (34) below] or of a test (8) [cf. eq. (33) below] is 1 irrespective of  $\alpha$  [the size of the test] if and only if  $f_{\max}$  is an eigenvector of  $W'$ .”

We note that, while not explicit in the above statement, it is understood in Martellosio (2010) that  $0 \leq \rho < \lambda_{\max}^{-1}$  is assumed. Similarly, the case  $n = 1$  is not ruled out explicitly in the statement of the proposition, but it seems to be implicitly understood in Martellosio (2010) that  $n \geq 2$  holds (note that in case  $n = 1$  the test statistics degenerate and therefore the associated tests trivially have size equal to 0 or 1, depending on the choice of the critical value).

The test defined in equation (8) of Martellosio (2010) rejects for small values of

$$y'(I_n - \bar{\rho}W')(I_n - \bar{\rho}W)y/\|y\|^2, \quad (33)$$

where  $0 < \bar{\rho} < \lambda_{\max}^{-1}$  is specified by the user. The argument in the proof of the proposition in Martellosio (2010) for this class of tests is incorrect for the following reasons:

1. The proof is based on Lemma D.3 in Martellosio (2010) which is incorrect as discussed in Appendix B.1.
2. Even if Lemma D.3 in Martellosio (2010) were correct and could be used, this lemma would only deliver the result  $\alpha^* = 0$  which does *not* imply, without a further argument, that the limiting power is equal to one for every size  $\alpha \in (0, 1)$ . By definition of  $\alpha^*$ ,  $\alpha^* = 0$  only implies that the limiting power is *nonzero* for every size  $\alpha \in (0, 1)$ .

For the case of the Cliff-Ord test, i.e., the test rejecting for small values of

$$-y'Wy/\|y\|^2 = -0.5y'(W + W')y/\|y\|^2, \quad (34)$$

Martellosio (2010) argues that this can be reduced to the previously considered case, the proof of which is flawed as just shown. Apart from this, the reduction argument, which we now quote, has its own problems:

”... By Lemma D.3 with  $B = \Gamma^{-1}(\bar{\rho})$  [which equals  $(I_n - \bar{\rho}W')(I_n - \bar{\rho}W)$ ], in order to prove that the limiting power of test (8) [cf. eq. (33) above] is 1 for any  $\alpha$  [the size of the test], we need to show that  $W'f_{\max} = \lambda_{\max}f_{\max}$  is necessary and sufficient for  $f_{\max} \in E_n(\Gamma(\bar{\rho}))$ . Clearly, if this holds for any  $\bar{\rho} > 0$ , it holds for  $\bar{\rho} \rightarrow 0$  too, establishing also the part of the proposition regarding the Cliff-Ord test. ...”

The problem here is that it is less than clear what the precise mathematical ”approximation” argument is. If we interpret it as deriving limiting power equal to 1 for the Cliff-Ord test from the corresponding result for tests of the form (8) and the fact that the Cliff-Ord test emerges as a limit of these tests for  $\bar{\rho} \rightarrow 0$ , then this involves an interchange of two limiting operations, namely  $\rho \rightarrow \lambda_{\max}^{-1}$  and  $\bar{\rho} \rightarrow 0$ , for which no justification is provided. Alternatively, one could try to interpret the ”approximation” argument as an argument that tries to derive  $f_{\max} \in E_n(W + W')$

from  $f_{\max} \in E_n(\Gamma(\bar{\rho}))$  for every  $\bar{\rho} > 0$ ; of course, such an argument would need some justification which, however, is not provided. We note that this argument could perhaps be saved by using the arguments we provide in the proof of Proposition 4.10, but the proof of our correct version of Proposition 1 in Martellosio (2010), i.e., Proposition 4.9 in Section 4.1, is more direct and does not need such a reasoning. Furthermore, note that the proof of Proposition 4.9 is based on our Proposition 2.26, which is a correct version of Lemma D.3 in Martellosio (2010) and which delivers not only the conclusion  $\alpha^* = 0$ , but the stronger conclusion that the limiting power is indeed equal to 1 for every size in  $(0, 1)$ .

We now turn to a discussion of Lemma E.4, which is again a statement about the Cliff-Ord test and tests of the form (8) in Martellosio (2010), but now in the context of the SEM (i.e.,  $k > 0$  is possible). The statement and the proof of the lemma suffer from the following shortcomings (again Lemma E.4 implicitly assumes that  $0 \leq \rho < \lambda_{\max}^{-1}$  holds):

1. The proof of the lemma is based on Lemma D.3 in Martellosio (2010), which is incorrect (cf. the discussion in Appendix B.1).
2. The proof uses non-rigorous arguments such as arguments involving a ‘limiting matrix’ with an infinite eigenvalue. Additionally, continuity of the dependence of eigenspaces on the underlying matrix is used without providing the necessary justification.
3. For the case of the Cliff-Ord test the same unjustified reduction argument as in the proof of Proposition 1 of Martellosio (2010) is used, cf. the preceding discussion.

For a correct version of Lemma E.4 of Martellosio (2010) see Proposition 4.10 in Section 4.1 above. As a point of interest we furthermore note that cases where the test statistics become degenerate (e.g., the case  $n - k = 1$ ) are not ruled out explicitly in Lemma E.4 in Martellosio (2010); in these cases  $\alpha^* = 1$  (and not  $\alpha^* = 0$ ) holds.

### B.3 Comments on Propositions 3, 4, and 5 in Martellosio (2010)

The proof of the part of Proposition 3 of Martellosio (2010) regarding point-optimal invariant tests seems to be correct except for the case where  $\text{span}(X)^\perp$  is contained in one of the eigenspaces of  $\Sigma(\rho)$ . In this case the test statistic of the form (8) in Martellosio (2010) is degenerate (see Section 4.3 above) and does not give the point-optimal invariant test (except in the trivial case where the size is 0 or 1, a case always excluded in Martellosio (2010)). However, this problem is easily fixed by observing that the point-optimal invariant test in this case is given by the randomized test  $\varphi \equiv \alpha$ , which is trivially unbiased. Two minor issues in the proof are as follows: (i) Lemma E.3 can only be applied as long as  $\mathbf{z}_i^2 > 0$  for every  $i \in H$ . Fortunately, the complement of this event is a null-set allowing the argument to go through. (ii) The expression ‘stochastically larger’ in the paragraph following (E.4) should read ‘stochastically smaller’. We also note that the assumption of Gaussianity can easily be relaxed to elliptical symmetry in view of  $G_X$ -invariance of the tests considered.

More importantly, the proof of the part of Proposition 3 of Martellosio (2010) concerning locally best invariant tests is highly deficient for at least two reasons: First, it is claimed that locally best invariant tests are of the form (7) in Martellosio (2010) with  $Q = d\Sigma(\rho)/d\rho|_{\rho=0}$ . While this is correct under regularity conditions (including a differentiability assumption on  $\Sigma(\rho)$ ), such conditions are, however, missing in Proposition 3 of Martellosio (2010). Also, the case where  $\text{span}(X)^\perp$  is contained in one of the eigenspaces of  $\Sigma(\rho)$  has to be treated separately, as then the locally best invariant

test is given by the randomized test  $\varphi \equiv \alpha$ . Second, the proof uses once more an unjustified approximation argument in an attempt to reduce the case of locally best invariant tests to the case of point-optimal invariant tests. It is not clear what the precise nature of the approximation argument is. Furthermore, even if the approximation argument could be somehow repaired to deliver unbiasedness of locally best invariant tests, it is less than clear that strict unbiasedness could be obtained this way as strict inequalities are not preserved by limiting operations.

We next turn to the part of Proposition 4 of Martellosio (2010) regarding point-optimal invariant tests.<sup>40</sup> As in the case of Proposition 3 discussed above, the case where  $\text{span}(X)^\perp$  is contained in one of the eigenspaces of  $\Sigma(\rho)$  has to be treated separately, and Gaussianity can be relaxed to elliptical symmetry. We note that the clause ‘if and only if’ in the last but one line of p. 185 of Martellosio (2010) should read ‘if’. We also note that the verification of the first displayed inequality on p. 186 of Martellosio (2010) could be shortened (using Lemma E.3 (more precisely, the more general result referred to in the proof of this lemma) with  $a_i = \lambda_i(W)/\tau_i(\rho)$ ,  $b_i = \tau_i^2(\bar{\rho})$ , and  $p_i = \mathbf{z}_i^2/\tau_i^2(\rho)$  to conclude that the first display on p. 186 holds almost surely, and furthermore that it holds almost surely with equality if and only if all  $b_i$  or all  $a_i$  are equal, which is equivalent to all  $\lambda_i(W)$  for  $i \in H$  being equal).

Again, the proof of the part of Proposition 4 of Martellosio (2010) concerning locally best invariant tests is deficient as it is based on the same unjustified approximation argument mentioned before.

We next turn to Proposition 5 of Martellosio (2010). In the last of the three cases considered in this proposition, both test statistics are degenerate and hence the power functions are trivially constant equal to 0 or 1 (a case ruled out in Martellosio (2010)). More importantly, the proof of Proposition 5 is severely flawed for several reasons, of which we only discuss a few: First, the proof makes use of Corollary 1 of Martellosio (2010), the proof of which is based on the incorrect Theorem 1 in Martellosio (2010); it also makes use of Lemma E.4 and Proposition 4 of Martellosio (2010) which are incorrect as discussed before. Second, even if these results used in the proof were correct as they stand, additional problems would arise: Lemma E.4 only delivers  $\alpha^* = 0$ , and not the stronger conclusion that the limiting power equals 1, as would be required in the proof. Furthermore, Proposition 4 has Gaussianity of the errors as a hypothesis, while such an assumption is missing in Proposition 5.

We conclude by mentioning that a correct version of the part of Proposition 5 of Martellosio (2010) concerning tests of the form (8) in Martellosio (2010) can probably be obtained by substituting our Corollary 4.5 and Proposition 4.10 for Corollary 1 and Lemma E.4 of Martellosio (2010) in the proof, but we have not checked the details. For the Cliff-Ord test this does not seem to work in the same way as the corresponding case of Proposition 4 of Martellosio (2010) is lacking a proof as discussed before.

## C Proofs for Section 2.2

**Proof of Lemma 2.5:** Let  $\rho_m$  be a sequence in  $[0, a)$  converging to  $a$  and let

$$\sum_{j=1}^{n-1} \lambda_j(\Sigma(\rho_m)) v_j(\rho_m) v_j(\rho_m)' + \lambda_n(\Sigma(\rho_m)) v_n(\rho_m) v_n(\rho_m)'$$

---

<sup>40</sup>While not explicit in the statement of this proposition, it is implicitly assumed that  $0 \leq \rho < \lambda_{\max}^{-1}$  holds.

be a spectral decomposition of  $\Sigma(\rho_m)$ , with  $v_j(\rho_m)$  ( $j = 1, \dots, n$ ) forming an orthonormal basis of eigenvectors of  $\Sigma(\rho_m)$  and  $\lambda_j(\Sigma(\rho_m))$  for  $j = 1, \dots, n$  denoting the corresponding eigenvalues ordered from smallest to largest and counted with their multiplicities. Because  $\Sigma^{-1}(a-)$  is rank-deficient by assumption, we must have  $\lambda_1(\Sigma^{-1}(\rho_m)) \rightarrow 0$ , or equivalently  $\lambda_n^{-1}(\Sigma(\rho_m)) \rightarrow 0$ . Because the kernel of  $\Sigma^{-1}(a-)$  has dimension one and because of positive definiteness of  $\Sigma(\rho_m)$  we can infer the existence of some  $0 < M < \infty$  such that  $0 < \lambda_j(\Sigma(\rho_m)) < M$  must hold for every  $j = 1, \dots, n-1$  and  $m \in \mathbb{N}$ . As a consequence, the sum from 1 to  $n-1$  in the previous display, after being premultiplied by  $\lambda_n^{-1}(\Sigma(\rho_m))$ , converges to zero for  $m \rightarrow \infty$ . It remains to show that  $v_n(\rho_m)v_n(\rho_m)' \rightarrow ee'$ . Let  $m'$  be an arbitrary subsequence of  $m$ . By norm-boundedness of the sequence  $v_n(\rho_m)$  there exists another subsequence  $m''$  along which  $v_n(\rho_m)$  converges to some normalized vector  $e^*$ , say. Clearly

$$\Sigma^{-1}(\rho_{m''})v_n(\rho_{m''}) = \lambda_n^{-1}(\Sigma(\rho_{m''}))v_n(\rho_{m''}).$$

The left hand side in the previous display now converges to  $\Sigma^{-1}(a-)e^*$  while the right hand side converges to zero. Therefore  $e^*$  is an element of the (one-dimensional) kernel of  $\Sigma^{-1}(a-)$ . Since  $e^*$  is normalized, we must have  $e^*e^{*'} = ee'$ . This proves the claim as the subsequence  $m'$  was arbitrary.  $\blacksquare$

**Lemma C.1.** *Let  $\mathbf{v}_m$  be a sequence of random  $n$ -vectors such that  $E(\mathbf{v}_m) = 0$  and  $E(\|\mathbf{v}_m\|^2) < \infty$  and let  $\Omega_m = E(\mathbf{v}_m\mathbf{v}_m')$ . If  $\Omega_m \rightarrow ee'$  as  $m \rightarrow \infty$  for some  $e \in \mathbb{R}^n$ , then the sequence  $\mathbf{v}_m$  is tight and the support of every weak accumulation point of the sequence of distributions of  $\mathbf{v}_m$  is a subset of  $\text{span}(e)$ . If, in addition, every weak accumulation point of the distributions of  $\mathbf{v}_m$  has no mass at the origin and if  $e$  is normalized, then the distribution of  $\mathcal{I}_{0,\zeta_e}(\mathbf{v}_m)$  converges weakly to  $\delta_e$ .*

**Proof:** Let  $M$  be an arbitrary positive real number. Since the sequence  $\text{trace}(\Omega_m)$  is convergent to  $\text{trace}(ee')$ , it is bounded from above by  $S$ , say. Markov's inequality gives

$$\Pr(\|\mathbf{v}_m\| \geq M) = \Pr(\|\mathbf{v}_m\|^2 \geq M^2) \leq \frac{E(\mathbf{v}_m'\mathbf{v}_m)}{M^2} = \frac{\text{trace}(\Omega_m)}{M^2} \leq \frac{S}{M^2}$$

for every  $m \in \mathbb{N}$ , which implies tightness. To prove the claim about the support of weak accumulation points note that  $\mathbf{v}_m = \Pi_{\text{span}(e)}\mathbf{v}_m + \Pi_{\text{span}(e)^\perp}\mathbf{v}_m$  and that the support of  $\Pi_{\text{span}(e)}\mathbf{v}_m$  is certainly a subset of  $\text{span}(e)$ , which is a closed set. It thus suffices to show that  $\Pi_{\text{span}(e)^\perp}\mathbf{v}_m$  converges to zero in probability. But this is again a consequence of Markov's inequality: For every  $\varepsilon > 0$  we have

$$\Pr(\|\Pi_{\text{span}(e)^\perp}\mathbf{v}_m\| \geq \varepsilon) \leq \frac{E(\|\Pi_{\text{span}(e)^\perp}\mathbf{v}_m\|^2)}{\varepsilon^2} = \frac{E(\mathbf{v}_m'\Pi_{\text{span}(e)^\perp}\mathbf{v}_m)}{\varepsilon^2} = \frac{\text{trace}(\Pi_{\text{span}(e)^\perp}\Omega_m)}{\varepsilon^2}. \quad (35)$$

Because  $\Omega_m \rightarrow ee'$ , we obtain  $\Pi_{\text{span}(e)^\perp}\Omega_m \rightarrow 0$  and hence the upper bound in (35) converges to zero as  $m \rightarrow \infty$ . To prove the final assertion let  $m'$  be an arbitrary subsequence and  $m''$  a subsequence thereof such that  $\mathbf{v}_{m''}$  converges weakly to  $\mathbf{v}$ , say. By what has already been established, we may assume that  $\Pi_{\text{span}(e)}\mathbf{v} = \mathbf{v}$  almost surely holds. Because  $\mathcal{I}_{0,\zeta_e}$  is continuous at  $\lambda e$  for every  $\lambda \neq 0$  and because  $\Pr(\mathbf{v} = 0) = 0$  by the assumptions, we can apply the continuous mapping theorem to conclude that

$$\mathcal{I}_{0,\zeta_e}(\mathbf{v}_{m''}) \rightarrow \mathcal{I}_{0,\zeta_e}(\mathbf{v}).$$

Since  $\Pr(\mathbf{v} = 0) = 0$ , we have that  $\mathcal{I}_{0,\zeta_e}(\mathbf{v})$  is almost surely equal to  $\zeta_e\left(\frac{\mathbf{v}}{\|\mathbf{v}\|}\right)$ . But this is almost surely equal to  $e$  by definition of  $\zeta_e$ . This completes the proof because  $m'$  was an arbitrary subsequence.  $\blacksquare$

**Proof of Proposition 2.6:** 1. Let  $\rho_m \in [0, a)$  be a sequence converging to  $a$ . Assumption 3 implies that  $P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}$  coincides with  $P_{0, \lambda_n^{-1/2}(\Sigma(\rho_m)), \rho_m}$ , which is precisely the distribution of  $\lambda_n^{-1/2}(\Sigma(\rho_m))\Sigma^{1/2}(\rho_m)\mathbf{z}$ . By Assumption 1 we have  $\lambda_n^{-1}(\Sigma(\rho_m))\Sigma(\rho_m) \rightarrow ee'$ . By continuity of the symmetric nonnegative definite square root we obtain

$$\lambda_n^{-1/2}(\Sigma(\rho_m))\Sigma^{1/2}(\rho_m) = (\lambda_n^{-1}(\Sigma(\rho_m))\Sigma(\rho_m))^{1/2} \rightarrow (ee')^{1/2} = ee'.$$

Consequently,  $\lambda_n^{-1/2}(\Sigma(\rho_m))\Sigma^{1/2}(\rho_m)\mathbf{z}$  converges weakly to  $ee'\mathbf{z}$ . Hence, the only accumulation point  $P$ , say, of  $P_{0, \lambda_n^{-1/2}(\Sigma(\rho_m)), \rho_m}$  is the distribution of  $ee'\mathbf{z}$ . The claim now follows because  $P(\{0\}) = \Pr(ee'\mathbf{z} = 0) = \Pr(e'\mathbf{z} = 0) = 0$  by assumption

2. Let  $\rho_m$  be as before and observe that again  $P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}$  coincides with  $P_{0, \lambda_n^{-1/2}(\Sigma(\rho_m)), \rho_m}$ , which, however, now equals the distribution of  $\lambda_n^{-1/2}(\Sigma(\rho_m))L(\rho_m)\mathbf{z}$ . Since  $L(\rho_m)$  is a square root of  $\Sigma(\rho_m)$ , there must exist an orthogonal matrix  $U(\rho_m)$  such that  $L(\rho_m) = \Sigma^{1/2}(\rho_m)U(\rho_m)$ . Rewrite  $\lambda_n^{-1/2}(\Sigma(\rho_m))L(\rho_m)$  as  $\lambda_n^{-1/2}(\Sigma(\rho_m))\Sigma^{1/2}(\rho_m)U(\rho_m)$ . Fix an arbitrary subsequence  $m'$  of  $m$ . Along a suitable subsubsequence  $m''$  the matrix  $U(\rho_{m''})$  converges to an orthogonal matrix  $U$ , say. Therefore  $\lambda_n^{-1/2}(\Sigma(\rho_{m''}))\Sigma^{1/2}(\rho_{m''})U(\rho_{m''})$  converges to  $ee'U$ . Hence, the only accumulation point  $P$ , say, of  $P_{0, \lambda_n^{-1/2}(\Sigma(\rho_m)), \rho_m}$  along the subsequence  $m''$  is the distribution of  $ee'U\mathbf{z}$ . But clearly  $P(\{0\}) = \Pr(ee'U\mathbf{z} = 0) = \Pr(e'U\mathbf{z} = 0)$ . Now this is equal to 0 in case the distribution of  $\mathbf{z}$  is dominated by  $\mu_{\mathbb{R}^n}$  since the set  $\{y \in \mathbb{R}^n : e'Uy = 0\}$  is obviously a  $\mu_{\mathbb{R}^n}$ -null set. Since  $m'$  was arbitrary, the proof of the first claim is complete. To prove the second claim observe that  $\Pr(e'U\mathbf{z} = 0) = \Pr(e'U(\mathbf{z}/\|\mathbf{z}\|) = 0)$ , which equals zero since the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  is dominated by  $\nu_{S^{n-1}}$  by assumption and since  $\{y \in S^{n-1} : e'Uy = 0\}$  is a  $\nu_{S^{n-1}}$ -null set (cf. Remark E.2(i)). ■

**Proof of Theorem 2.7:** Let  $\rho_m$  be a sequence in  $[0, a)$  converging to  $a$ . Invariance of the test  $\varphi$  w.r.t.  $G_X$  implies

$$\begin{aligned} E_{\beta, \sigma, \rho_m}(\varphi) &= \int_{\mathbb{R}^n} \varphi(y) dP_{\beta, \sigma, \rho_m} = \int_{\mathbb{R}^n} \varphi\left(M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}(y)\right) dP_{\beta, \sigma, \rho_m} \\ &= \int_{\mathbb{R}^n} \varphi(y) d\left(P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}\right) \\ &= \int_{\mathbb{R}^n} \varphi(\mathcal{I}_{0, \zeta_e}(y)) d\left(P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}\right) \\ &= \int_{\mathbb{R}^n} \varphi(y) d\left(\left(P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}\right) \circ \mathcal{I}_{0, \zeta_e}\right), \end{aligned}$$

where the last but one equality holds because of Remark 2.2(ii). The covariance matrix of  $\mathbf{v}_m$ , say, a centered random variable with distribution  $P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}$ , is given by  $\lambda_n^{-1}(\Sigma(\rho_m))\Sigma(\rho_m)$  which converges to  $ee'$  by Assumption 1. Note that  $e$  is necessarily normalized. By Assumption 2 every weak accumulation point  $P$  of  $P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}$  satisfies  $P(\{0\}) = 0$  (note that  $P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}$  is in fact tight by Lemma C.1). Thus we can apply Lemma C.1 to conclude that

$$\left(P_{\beta, \sigma, \rho_m} \circ M_{X\beta, \lambda_n^{1/2}(\Sigma(\rho_m))\sigma}\right) \circ \mathcal{I}_{0, \zeta_e} \rightarrow \delta_e$$

weakly as  $m \rightarrow \infty$ . Since  $\varphi$  is bounded and is continuous at  $e$ , the claim then follows from a version of the Portmanteau theorem, cf. Theorem 30.12 in Bauer (2001). ■

**Proof of Proposition 2.11:** 1. Because  $\emptyset \neq \Phi \neq \mathbb{R}^n$  we can find  $y_0 \in \Phi$  and  $y_1 \notin \Phi$ . By  $G_X$ -invariance we have that  $\gamma y_0 + X\theta \in \Phi$  and  $\gamma y_1 + X\theta \notin \Phi$  for every  $\gamma \neq 0$  and for every  $\theta \in \mathbb{R}^k$ . Letting  $\gamma$  converge to zero we see that  $X\theta$  belongs to the closure of  $\Phi$  as well as of its complement. Thus  $X\theta \in \text{bd}(\Phi)$  holds for every  $\theta$ .

2. Suppose  $y$  is an element of the boundary of the rejection region. If  $y \in \text{span}(X)$  there is nothing to prove. Hence assume  $y \notin \text{span}(X)$ . If  $T(y) \neq \kappa$  would hold, then by the continuity assumption  $y$  would be either in the interior or the exterior (i.e., the complement of the closure) of the rejection region.

3. Because  $T_B$  is continuous on  $\mathbb{R}^n \setminus \text{span}(X)$ , Part 2 of the proposition establishes that the l.h.s. of (9) is contained in the r.h.s. Because of Part 1, it suffices to show that every  $y_0 \notin \text{span}(X)$  satisfying  $T_B(y_0) = \kappa$  belongs to  $\text{bd}(\Phi_{B,\kappa})$ . Obviously,  $y_0 \notin \Phi_{B,\kappa}$ . It remains to show that  $y_0$  can be approximated by a sequence of elements belonging to  $\Phi_{B,\kappa}$ : For  $\lambda \in \mathbb{R}$  set  $y(\lambda) = y_0 + \lambda y_*$  where  $y_* \in \mathbb{R}^n$  is such that  $T_B(y_*) > \kappa$ . Such an  $y_*$  exists, because  $\Phi_{B,\kappa} \neq \emptyset$  by assumption. Furthermore,  $y_* \notin \text{span}(X)$  must hold, since otherwise  $\lambda_1(B) = T_B(y_*) > \kappa$  would follow, which in turn would entail  $T_B(y) \geq \lambda_1(B) > \kappa$  for all  $y \in \mathbb{R}^n$ , i.e.,  $\Phi_{B,\kappa} = \mathbb{R}^n$ , contradicting the assumptions. Set  $A = C'_X(B - \kappa I_{n-k})C_X$  and note that  $y'_0 A y_0 = 0$  and  $y'_* A y_* > 0$  hold. Now

$$y(\lambda)' A y(\lambda) = y'_0 A y_0 + 2\lambda y'_0 A y_* + \lambda^2 y'_* A y_* = 2\lambda y'_0 A y_* + \lambda^2 y'_* A y_*.$$

Choose a sequence  $\lambda_m$  that converges to zero for  $m \rightarrow \infty$  and satisfies  $\lambda_m > 0$  for all  $m$  if  $y'_0 A y_* \geq 0$  and  $\lambda_m < 0$  for all  $m$  if  $y'_0 A y_* < 0$ . Then  $y(\lambda_m)$  converges to  $y_0$  and  $y(\lambda_m) \notin \text{span}(X)$  holds for large enough  $m$ . Furthermore, we have  $y'(\lambda_m) A y(\lambda_m) > 0$ . But this means that  $T_B(y(\lambda_m)) > \kappa$  holds for large  $m$ . ■

**Proof of Lemma 2.14:** Suppose Assumption 4 holds. Then clearly

$$\lim_{\rho \rightarrow a} c^2(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)^\perp} = \Lambda \Lambda' \quad (36)$$

holds. Set  $V = \Lambda \Lambda'$ . Furthermore, the above relation clearly implies  $V e = 0$  and hence  $\text{span}(e) \subseteq \ker(V)$ . Because  $V y = 0$  if and only if  $\Lambda' y = 0$ , and because  $\text{rank}(\Lambda') = \text{rank}(\Lambda) = n - 1$ , it follows that  $\ker(V)$  must be one-dimensional. Hence  $\ker(V) = \text{span}(e)$  must hold. Since  $V$  maps  $\text{span}(e)^\perp$  into  $\text{span}(e)^\perp$  in view of (36), it follows that  $V$  is injective on  $\text{span}(e)^\perp$ . To prove the converse, note that  $V$  given by (10) is by construction a bijection from  $\text{span}(e)^\perp$  to itself and is symmetric and nonnegative definite. Thus its symmetric nonnegative definite square root  $V^{1/2}$  exists and is a bijective map from  $\text{span}(e)^\perp$  to itself. Furthermore, the symmetric nonnegative square root of  $c^2(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)^\perp}$  can be written in the form  $c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma^{1/2}(\rho) U(\rho)$  for a suitable choice of an orthogonal matrix  $U(\rho)$ . By continuity of the symmetric nonnegative square root we obtain

$$c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma^{1/2}(\rho) U(\rho) \rightarrow V^{1/2}.$$

It remains to set  $\Lambda = V^{1/2}$  and  $L_*(\rho) = \Sigma^{1/2}(\rho) U(\rho)$ . ■

**Proof of Theorem 2.16:** A.1. By  $G_X$ -invariance of  $\varphi$  and Assumption 3 the power function does neither depend on  $\beta$  nor  $\sigma$  (cf. Remark 2.3), and thus it suffices to consider the case  $\beta = 0$  and  $\sigma = 1$ . By Assumption 3 we furthermore have

$$E_{0,1,\rho}(\varphi) = \int_{\mathbb{R}^n} \varphi dP_{0,1,\rho} = \int \varphi(L(\rho)\mathbf{z}) d\text{Pr} = \int \varphi(L_*(\rho)U(\rho)\mathbf{z}) d\text{Pr} \quad (37)$$

where  $U(\rho) = L_*^{-1}(\rho)L(\rho)$  is an orthogonal matrix. Observe that  $\varphi(y + \gamma e) = \varphi(y)$  holds for every  $y$  and for every  $\gamma \in \mathbb{R}$ : This is trivial for  $\gamma = 0$  and follows for  $\gamma \neq 0$  from

$$\varphi(y + \gamma e) = \varphi(\gamma^{-1}y + e) = \varphi(\gamma^{-1}y) = \varphi(y), \quad (38)$$

where we have made use of  $G_X$ -invariance of  $\varphi$  as well as of (11). Observing that  $\Pi_{\text{span}(e)}L_*(\rho)U(\rho)\mathbf{z}$  as well as  $\Pi_{\text{span}(e)}U(\rho)\mathbf{z}$  belong to  $\text{span}(e)$ , using relation (38) as well as  $G_X$ -invariance of  $\varphi$  leads to

$$\begin{aligned} \int \varphi(L_*(\rho)U(\rho)\mathbf{z})d\Pr &= \int \varphi(\Pi_{\text{span}(e)^\perp}L_*(\rho)U(\rho)\mathbf{z})d\Pr \\ &= \int \varphi(c(\rho)\Pi_{\text{span}(e)^\perp}L_*(\rho)U(\rho)\mathbf{z})d\Pr \\ &= \int \varphi(A(\rho)U(\rho)\mathbf{z})d\Pr, \end{aligned} \quad (39)$$

where  $A(\rho)$  is shorthand for  $\Pi_{\text{span}(e)} + c(\rho)\Pi_{\text{span}(e)^\perp}L_*(\rho)$ . Since the image of  $\Lambda$  is  $\text{span}(e)^\perp$  and  $\Lambda$  is injective when restricted to  $\text{span}(e)^\perp$  it follows that  $A := \Pi_{\text{span}(e)} + \Lambda$  is bijective as a map from  $\mathbb{R}^n$  to  $\mathbb{R}^n$ . [To see this suppose that  $Ay = 0$ . Because  $\Lambda y \in \text{span}(e)^\perp$  this implies  $\Pi_{\text{span}(e)}y = 0$  as well as  $\Lambda y = 0$ . The first equality now implies  $y \in \text{span}(e)^\perp$ . Bijectivity of  $\Lambda$  on  $\text{span}(e)^\perp$  then implies  $y = 0$ .] By Assumption 4 the matrix  $A(\rho)$  converges to  $A$  for  $\rho \rightarrow a$  and thus  $A(\rho)$  is bijective as a map from  $\mathbb{R}^n$  to  $\mathbb{R}^n$  whenever  $\rho$  is sufficiently close to  $a$ , say  $\rho \geq \rho_0$ . If now  $\omega$  is an accumulation point of  $E_{0,1,\rho}(\varphi)$ , we can find a sequence  $\rho_m$  that converges to  $a$  such that  $E_{0,1,\rho_m}(\varphi)$  converges to  $\omega$ . By passing to a suitable subsequence, we may also assume that  $U(\rho_m)$  converges to an orthogonal matrix  $U$ , say. W.l.o.g. we may furthermore assume that  $\rho_m \geq \rho_0$  holds and thus  $A(\rho_m)$  is nonsingular. By the transformation formula for densities the  $\mu_{\mathbb{R}^n}$ -density of the random vector  $A(\rho_m)U(\rho_m)\mathbf{z}$  is given by

$$|\det(A^{-1}(\rho_m))|p(U'(\rho_m)A^{-1}(\rho_m)y).$$

Because of  $A(\rho_m) \rightarrow A$ ,  $U(\rho_m) \rightarrow U$ , and because  $p$  is continuous  $\mu_{\mathbb{R}^n}$ -almost everywhere, this expression converges for  $\mu_{\mathbb{R}^n}$ -almost every  $y \in \mathbb{R}^n$  to

$$|\det(A^{-1})|p(U'A^{-1}y) \quad (40)$$

as  $\rho \rightarrow a$ , which is the density of the random vector  $AU\mathbf{z}$ . Scheffé's lemma thus implies that the distribution of  $A(\rho_m)U(\rho_m)\mathbf{z}$  converges in total variation norm to  $Q_{A,U}$ , the distribution of  $AU\mathbf{z}$ . It now follows in view of (37) and (39) that

$$E_{0,1,\rho_m}(\varphi) \rightarrow E_{Q_{A,U}}(\varphi) = \int \varphi dQ_{A,U} = \int \varphi(AU\mathbf{z})d\Pr.$$

Now

$$\varphi(AU\mathbf{z}) = \varphi(\Pi_{\text{span}(e)}U\mathbf{z} + \Lambda U\mathbf{z}) = \varphi(\Lambda U\mathbf{z})$$

holds because of (38), implying  $E_{Q_{A,U}}(\varphi) = E_{Q_{\Lambda,U}}(\varphi)$ . This shows that  $\omega = E_{Q_{\Lambda,U}}(\varphi)$  must hold. Conversely, given  $U \in \mathcal{U}(L_*^{-1}L)$  we can find a sequence  $\rho_m \rightarrow a$  such that  $U(\rho_m) = L_*^{-1}(\rho_m)L(\rho_m)$  converges to the given  $U$ . Repeating the argument given above then shows that  $E_{Q_{\Lambda,U}}(\varphi)$ , for the given  $U$ , arises as an accumulation point of  $E_{0,1,\rho}(\varphi)$  for  $\rho \rightarrow a$ .

A.2. The claim follows immediately from the already established Part 1.

A.3. Recall that  $E_{Q_{\Lambda,U}}(\varphi) = E_{Q_{A,U}}(\varphi)$ . Hence,

$$E_{Q_{\Lambda,U}}(\varphi) = \int_{\mathbb{R}^n} \varphi(AUy) p(y) dy = \int_{\mathbb{R}^n \setminus \{0\}} \varphi(AUy) p(y) dy = \int_{(0,\infty) \times S^{n-1}} \varphi(rAU s) p(rs) dH(r, s)$$

where  $H$  is the pushforward measure of  $\mu_{\mathbb{R}^n}$  (restricted to  $\mathbb{R}^n \setminus \{0\}$ ) under the map  $y \mapsto (\|y\|, y/\|y\|)$ . Now  $H$  is nothing else than the product of the measure on  $(0, \infty)$  with density  $r^{n-1}$  and the surface measure  $c\nu_{S^{n-1}}$  on  $S^{n-1}$  with the constant  $c$  given by  $2\pi^{n/2}/\Gamma(n/2)$  (cf. Stroock (1999)). In view of Fubini's theorem (observe all functions involved are nonnegative) and invariance of  $\varphi$  we then obtain

$$E_{Q_{\Lambda,U}}(\varphi) = c \int_{S^{n-1}} \varphi(AUs) \left( \int_{(0,\infty)} p_s(r) r^{n-1} dr \right) d\nu_{S^{n-1}}.$$

If  $\varphi(\cdot)$  is not equal to zero  $\mu_{\mathbb{R}^n}$ -almost everywhere, then so is  $\varphi(AU\cdot)$  because  $AU$  is nonsingular. Now scale invariance of  $\varphi$  translates into scale invariance of  $\varphi(AU\cdot)$ , and hence  $\varphi(AU\cdot)$  restricted to  $S^{n-1}$  is not equal to zero  $\nu_{S^{n-1}}$ -almost everywhere, cf. Remark E.2(i) in Appendix E. Since the inner integral in the preceding display is positive  $\nu_{S^{n-1}}$ -almost everywhere by the assumption on  $p$ , we conclude that  $E_{Q_{\Lambda,U}}(\varphi)$  must be positive. The claim that  $E_{Q_{\Lambda,U}}(\varphi) < 1$  is proved by applying the above to  $1 - \varphi$ . Hence, if  $\varphi$  is neither  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to zero nor  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to one, we have established that  $E_{Q_{\Lambda,U}}(\varphi)$  is strictly between 0 and 1. Next observe that  $\mathcal{U}(L_*^{-1}L)$  is a compact set. It thus suffices to establish that the map  $U \rightarrow E_{Q_{\Lambda,U}}(\varphi)$  is continuous on  $\mathcal{U}(L_*^{-1}L)$ . But this follows from (40),  $\mu_{\mathbb{R}^n}$ -almost sure continuity of  $p$ , and Scheffé's Lemma.

B. By the assumptions on  $\mathfrak{P}$  the random vector  $\mathbf{z}$  is spherically symmetric with  $\Pr(\mathbf{z} = 0) = 0$ , and hence is almost surely equal to  $\mathbf{r}\mathbf{E}$  where  $\mathbf{r} = \|\mathbf{z}\|$  is a random variable satisfying  $\Pr(\mathbf{r} > 0) = 1$  and where  $\mathbf{E} = \mathbf{z}/\|\mathbf{z}\|$  is independent of  $\mathbf{r}$  and is uniformly distributed on the unit sphere  $S^{n-1}$  (cf. Lemma 1 in Cambanis et al. (1981)). Possibly after enlarging the underlying probability space we can find a random variable  $\mathbf{r}_0$  which is independent of  $\mathbf{E}$  and which is distributed as the square root of a chi-square with  $n$  degrees of freedom. By  $G_X$ -invariance of  $\varphi$  we have

$$\begin{aligned} E_{0,1,\rho}(\varphi) &= \int \varphi(L(\rho)\mathbf{z}) d\Pr = \int \varphi(L(\rho)\mathbf{r}\mathbf{E}) d\Pr = \int \varphi(L(\rho)\mathbf{E}) d\Pr \\ &= \int \varphi(L(\rho)\mathbf{r}_0\mathbf{E}) d\Pr = \int \varphi(L(\rho)\mathbf{G}) d\Pr \end{aligned} \quad (41)$$

where  $\mathbf{G} = \mathbf{r}_0\mathbf{E}$  has a standard multivariate Gaussian distribution. Again using  $G_X$ -invariance of  $\varphi$  we similarly obtain

$$E_{Q_{\Lambda,U}}(\varphi) = E\varphi(\Lambda U\mathbf{z}) = E\varphi(\Lambda U\mathbf{r}\mathbf{E}) = E\varphi(\Lambda U\mathbf{E}) = E\varphi(\Lambda U\mathbf{r}_0\mathbf{E}) = E\varphi(\Lambda U\mathbf{G}) = E_{Q_{\Lambda,U}^0}(\varphi)$$

where  $Q_{\Lambda,U}^0$  denotes the distribution of  $\Lambda U\mathbf{G}$ . This shows that we may act as if  $\mathbf{z}$  were Gaussian. Consequently, the results in A.1-A.3 apply. Furthermore, under elliptical symmetry  $Q_{\Lambda,U} = Q_{\Lambda,I_n}$  holds for every orthogonal matrix  $U$ . Hence, there exists only one accumulation point which is given by  $E_{Q_{\Lambda,I_n}}(\varphi)$ . [Alternatively, under elliptical symmetry we may choose w.l.o.g.  $L(\cdot)$  to be any square root of  $\Sigma(\cdot)$ , and thus equal to  $L_*(\cdot)$ , and then apply Part A.2.] That  $0 < E_{Q_{\Lambda,I_n}}(\varphi) < 1$  holds under the additional assumption on  $\varphi$  follows from Part A.3. ■



**Lemma C.2.** *Suppose Assumptions 1 and 4 hold with the same vector  $e$ . Then*

$$\Xi(\rho) := \lambda_n^{-1/2}(\Sigma(\rho)) c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho) \Pi_{\text{span}(e)}$$

*is bounded for  $\rho \rightarrow a$  and the set of all accumulation points of  $\Xi(\rho)$  for  $\rho \rightarrow a$  is given by*

$$\left\{ \Lambda U'_0 e e' : U_0 \in \mathcal{U} \left( \Sigma^{-1/2} L_* \right) \right\}.$$

*The same statements hold if  $\Xi(\rho)$  is replaced by  $\Xi_1(\rho) := \lambda_n^{-1/2}(\Sigma(\rho)) c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma(\rho)$  or  $\Xi_2(\rho) := c(\rho) \Pi_{\text{span}(e)^\perp} \Sigma^{1/2}(\rho) \Pi_{\text{span}(e)}$ .*

**Proof:** Rewrite  $\Xi(\rho)$  as  $A_1(\rho) U'(\rho) A_2(\rho) \Pi_{\text{span}(e)}$  where  $A_1(\rho) = c(\rho) \Pi_{\text{span}(e)^\perp} L_*(\rho)$ ,  $U(\rho) = \Sigma^{-1/2}(\rho) L_*(\rho)$  is orthogonal, and  $A_2(\rho) = \lambda_n^{-1/2}(\Sigma(\rho)) \Sigma^{1/2}(\rho)$ . Now  $A_1(\rho)$  and  $A_2(\rho)$  converge to  $\Lambda$  and  $ee'$ , respectively, by Assumptions 1 and 4. Since  $U(\rho)$  is clearly bounded, boundedness of  $\Xi(\rho)$  follows. The claim concerning the set of accumulation points also now follows immediately. The proofs for  $\Xi_1$  and  $\Xi_2$  are completely analogous. ■

**Proof of Theorem 2.18:** 1. Using invariance w.r.t.  $G_X$ , Equation (12), and homogeneity of  $D$  we obtain for every  $\gamma \neq 0$

$$T(\gamma e + h) = T(e + \gamma^{-1} h) = T(e) + \gamma^{-q} D(h) + R(\gamma^{-1} h) \quad (42)$$

for every  $h \in \mathbb{R}^n$ . Let  $\omega$  be an accumulation point of  $P_{\beta, \sigma, \rho}(\{y \in \mathbb{R}^n : T(y) > T(e)\})$  for  $\rho \rightarrow a$ . Then we can find a sequence  $\rho_m \in [0, a)$  with  $\rho_m \rightarrow a$  along which the rejection probability converges to  $\omega$ . W.l.o.g. (possibly after passing to a suitable subsequence) we may also assume that along this sequence the orthogonal matrices  $U(\rho_m) = L_*^{-1}(\rho_m) L(\rho_m)$  and  $U_0(\rho_m) = \Sigma^{-1/2}(\rho_m) L(\rho_m)$  converge to orthogonal matrices  $U$  and  $U_0$ , respectively. Using  $\Pi_{\text{span}(e)} = ee'$  and invariance w.r.t.  $G_X$  we obtain

$$T(X\beta + \sigma L(\rho_m)\mathbf{z}) = T(ee' L(\rho_m)\mathbf{z} + \Pi_{\text{span}(e)^\perp} L(\rho_m)\mathbf{z}).$$

Observe that  $ee' L(\rho_m)\mathbf{z}$  is nonzero with probability 1 because  $e \neq 0$ ,  $L(\rho_m)$  is nonsingular, and  $\mathbf{z}$  possesses a density. Hence, combining the previous display and equation (42) with  $\gamma = ee' L(\rho_m)\mathbf{z}$  and  $h = \Pi_{\text{span}(e)^\perp} L(\rho_m)\mathbf{z}$  and then multiplying by  $c^q(\rho_m) \lambda^{q/2}(m)$ , where  $\lambda(m)$  is shorthand for  $\lambda_n(\Sigma(\rho_m))$ , we obtain that

$$\begin{aligned} & c^q(\rho_m) \lambda^{q/2}(m) (T(X\beta + \sigma L(\rho_m)\mathbf{z}) - T(e)) \\ &= \left( \lambda^{-1/2}(m) ee' L(\rho_m)\mathbf{z} \right)^{-q} D(c(\rho_m) \Pi_{\text{span}(e)^\perp} L_*(\rho_m) U(\rho_m)\mathbf{z}) \\ & \quad + c^q(\rho_m) \lambda^{q/2}(m) R \left( \left( \lambda^{-1/2}(m) ee' L(\rho_m)\mathbf{z} \right)^{-1} \lambda^{-1/2}(m) \Pi_{\text{span}(e)^\perp} L(\rho_m)\mathbf{z} \right) \end{aligned} \quad (43)$$

holds almost surely. Next observe that by Assumption 1, continuity of the symmetric nonnegative definite square root, and  $(ee')^{1/2} = ee'$  we have

$$\lambda^{-1/2}(m) ee' L(\rho_m)\mathbf{z} = \lambda^{-1/2}(m) ee' \Sigma^{1/2}(\rho_m) U_0(\rho_m)\mathbf{z} \rightarrow ee' U_0 \mathbf{z} \quad (44)$$

and

$$\lambda^{-1/2}(m) \Pi_{\text{span}(e)^\perp} L(\rho_m)\mathbf{z} = \lambda^{-1/2}(m) \Pi_{\text{span}(e)^\perp} \Sigma^{1/2}(\rho_m) U_0(\rho_m)\mathbf{z} \rightarrow \Pi_{\text{span}(e)^\perp} ee' U_0 \mathbf{z} = 0, \quad (45)$$

where the convergence holds for every realization of  $\mathbf{z}$ . Note that  $e'U_0\mathbf{z} \neq 0$  holds almost surely. Relation (44) together with Assumption 4 then implies that the first term on the r.h.s. of (43) converges almost surely to  $(e'U_0\mathbf{z})^{-q} D(\Lambda U\mathbf{z})$  since  $D$  is clearly continuous. We next show that the second term on the r.h.s. of (43) converges to zero almost surely: Let  $h_m$  denote the argument of  $R$  in (43). Fix a realization of  $\mathbf{z}$  such that  $e'U_0\mathbf{z} \neq 0$ . Then  $h_m$  is well-defined for large enough  $m$ , and it converges to zero because of (44) and (45). Since  $R(0) = 0$  holds as a consequence of (12), we only need to consider subsequences along which  $h_m \neq 0$ . For notational convenience we denote such subsequences again by  $h_m$ . Because of the assumptions on  $R$  it suffices to show that  $c^q(\rho_m)\lambda^{q/2}(m)\|h_m\|^q$  is bounded. Now

$$\begin{aligned} & c^q(\rho_m)\lambda^{q/2}(m)\|h_m\|^q \\ &= \left\| \left( \lambda^{-1/2}(m)e' L(\rho_m)\mathbf{z} \right)^{-1} c(\rho_m)\Pi_{\text{span}(e)^\perp} L(\rho_m)\mathbf{z} \right\|^q \\ &= \left\| \left( \lambda^{-1/2}(m)e' L(\rho_m)\mathbf{z} \right)^{-1} c(\rho_m)\Pi_{\text{span}(e)^\perp} L_*(\rho_m)U(\rho_m)\mathbf{z} \right\|^q \\ &\rightarrow \left\| (e'U_0\mathbf{z})^{-1} \Lambda U\mathbf{z} \right\|^q < \infty, \end{aligned}$$

where we have made use of (44) and Assumption 4. We have thus established that

$$c^q(\rho_m)\lambda^{q/2}(m)(T(X\beta + \sigma L(\rho_m)\mathbf{z}) - T(e)) \rightarrow (e'U_0\mathbf{z})^{-q} D(\Lambda U\mathbf{z}) \quad (46)$$

almost surely. Note that the range of  $\Lambda$  is  $\text{span}(e)^\perp$ , and that  $\Lambda$  is bijective as a map from  $\text{span}(e)^\perp$  to itself. Hence, the random variable  $\Lambda U\mathbf{z}$  takes its values in  $\text{span}(e)^\perp$  and possesses a density on this subspace (w.r.t.  $n-1$  dimensional Lebesgue measure on this subspace). Since  $D$  restricted to  $\text{span}(e)^\perp$  can be expressed as a multivariate polynomial (in  $n-1$  variables) and does not vanish identically on  $\text{span}(e)^\perp$ , it vanishes at most on a subset of  $\text{span}(e)^\perp$  that has  $(n-1)$ -dimensional Lebesgue measure zero. It follows that  $D(\Lambda U\mathbf{z})$ , and hence the limit in (46), is nonzero almost surely. Observe that

$$\begin{aligned} P_{\beta, \sigma, \rho_m}(\{y \in \mathbb{R}^n : T(y) > T(e)\}) &= \Pr(T(X\beta + \sigma L(\rho_m)\mathbf{z}) > T(e)) \\ &= \Pr\left(c^q(\rho_m)\lambda^{q/2}(m)(T(X\beta + \sigma L(\rho_m)\mathbf{z}) - T(e)) > 0\right) \end{aligned}$$

since  $c(\rho_m)$  and  $\lambda(m)$  are positive. By an application of the Portmanteau theorem we can thus conclude from (46) that for  $m \rightarrow \infty$

$$P_{\beta, \sigma, \rho_m}(\{y \in \mathbb{R}^n : T(y) > T(e)\}) \rightarrow \Pr((e'U_0\mathbf{z})^{-q} D(\Lambda U\mathbf{z}) > 0). \quad (47)$$

The limit in the preceding display obviously reduces to (14) and (15), respectively, and clearly  $(U, U_0) \in \mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  implies  $U \in \mathcal{U}(L_*^{-1}L)$ . This together then proves that every accumulation point  $\omega$  has the claimed form. To prove the converse, observe first that for every  $U \in \mathcal{U}(L_*^{-1}L)$  we can find an  $U_0$  such that  $(U, U_0) \in \mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  holds (exploiting compactness of the set of orthogonal matrices). Now, let  $(U, U_0) \in \mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  be given. Then we can find a sequence  $\rho_m \in [0, a]$  with  $\rho_m \rightarrow a$  such that  $U(\rho_m) = L_*^{-1}(\rho_m)L(\rho_m)$  and  $U_0(\rho_m) = \Sigma^{-1/2}(\rho_m)L(\rho_m)$  converge to  $U$  and  $U_0$ , respectively. Repeating the preceding arguments, then shows that  $\Pr((e'U_0\mathbf{z})^{-q} D(\Lambda U\mathbf{z}) > 0)$  is the limit of  $P_{\beta, \sigma, \rho_m}(\{y \in \mathbb{R}^n : T(y) > T(e)\})$ . The final claim is now obvious.

2. If  $\mathfrak{P}$  is an elliptically symmetric family we can w.l.o.g. set  $L(\cdot) = L_*(\cdot)$ , implying that  $\mathcal{U}(L_*^{-1}L, \Sigma^{-1/2}L)$  reduces to  $\{I_n\} \times \mathcal{U}(\Sigma^{-1/2}L_*)$ . Furthermore, as  $\mathbf{z}$  is then spherically symmetric and satisfies  $\Pr(\mathbf{z} = 0) = 0$ , it is almost surely equal to  $\mathbf{r}\mathbf{E}$  where  $\mathbf{r}$  must satisfy  $\Pr(\mathbf{r} > 0) = 1$  and where  $\mathbf{E}$  is independent of  $\mathbf{r}$  and is uniformly distributed on the unit sphere in  $\mathbb{R}^n$ . Let  $\mathbf{r}_0$  be a random variable which is independent of  $\mathbf{E}$  and which is distributed as the square root of a chi-square with  $n$  degrees of freedom (this may require enlarging the underlying probability space) and define  $\mathbf{G} = \mathbf{r}_0\mathbf{E}$  which clearly is a multivariate Gaussian random vector with mean zero and covariance matrix  $I_n$ . Define  $\mathfrak{P}_0$  in the same way as  $\mathfrak{P}$ , but with  $\mathbf{G}$  replacing  $\mathbf{z}$  in Assumption 3. Observe that the rejection probabilities of the test considered are the same whether they are calculated under the experiment  $\mathfrak{P}$  or  $\mathfrak{P}_0$  because of  $G_X$ -invariance of the test statistic. Applying the already established Part 1 in the context of the experiment  $\mathfrak{P}_0$  thus shows that the accumulation points of the rejection probabilities calculated under  $\mathfrak{P}_0$  as well as under  $\mathfrak{P}$  equal  $\Pr(D(\Lambda\mathbf{G}) > 0)$  for even  $q$  and equal  $\Pr(D(\Lambda\mathbf{G}) > 0, e'U_0\mathbf{G} > 0) + \Pr(D(\Lambda\mathbf{G}) < 0, e'U_0\mathbf{G} < 0)$  for odd  $q$ . In view of homogeneity of  $D$  and the fact that  $\mathbf{r}$  as well as  $\mathbf{r}_0$  are almost surely positive, these probabilities do not change their value if we replace  $\mathbf{G}$  by  $\mathbf{z}$ . This proves (16) and (17). To prove the last but one claim observe that  $E((e'U_0\mathbf{G})\Lambda\mathbf{G}) = \Lambda U'_0e = 0$ . Consequently,  $e'U_0\mathbf{G}$  and  $\Lambda\mathbf{G}$  are independent. Hence the accumulation point can be written as

$$\Pr(D(\Lambda\mathbf{G}) > 0)\Pr(e'U_0\mathbf{G} > 0) + \Pr(D(\Lambda\mathbf{G}) < 0)\Pr(e'U_0\mathbf{G} < 0).$$

This reduces to  $1/2$ , because then obviously  $\Pr(e'U_0\mathbf{G} > 0) = \Pr(e'U_0\mathbf{G} < 0) = 1/2$  (note that  $\Pr(e'U_0\mathbf{G} = 0) = 0$ ) and because  $\Pr(D(\Lambda\mathbf{G}) = 0) = 0$  (which is proved by arguments similar to the ones given below (46)). The final claim follows because by the assumed symmetry  $\Lambda U'_0e = U_0\Lambda'e = 0$ , the last equality following from the definition of  $\Lambda$ .

3. Lemma C.2 shows that under the additional assumption we have  $\Lambda U'_0ee' = 0$  for every  $U_0 \in \mathcal{U}(\Sigma^{-1/2}L_*)$ , and hence  $\Lambda U'_0e = 0$ . The claim then follows from Part 2. ■

**Lemma C.3.** *Suppose  $T$  is a test statistic that satisfies the conditions imposed on  $T$  in Theorem 2.18 for some normalized vector  $e$ . Then:*

1.  $D(h) = D(\Pi_{\text{span}(e)^\perp}h)$  holds for every  $h \in \mathbb{R}^n$ . In particular,  $D$  vanishes on all of  $\text{span}(e)$ .
2. If  $D(h) < 0$  holds for every  $h \in \text{span}(e)^\perp$  with  $h \neq 0$ , then there exists a neighborhood of  $e$  in  $\mathbb{R}^n$  such that  $T(y) \leq T(e)$  holds for every  $y$  in that neighborhood.

**Proof:** 1. Write  $h$  as  $\gamma e + h_2$  with  $h_2 = \Pi_{\text{span}(e)^\perp}h$ . Then for every sufficiently small real  $c > 0$  we have  $1 + c\gamma \neq 0$ , and hence exploiting  $G_X$ -invariance of  $T$  we obtain

$$T(e + ch) = T((1 + c\gamma)e + ch_2) = T\left(e + (1 + c\gamma)^{-1}ch_2\right).$$

Applying (12) to both sides of the above equation, using homogeneity of  $D$ , and dividing by  $c^{-q}$  we arrive at

$$D(h) + c^{-q}R(ch) = (1 + c\gamma)^{-q}D(h_2) + c^{-q}R\left((1 + c\gamma)^{-1}ch_2\right).$$

Now observe that  $c^{-q}R(ch)$  is zero for  $h = 0$ , and converges to zero for  $c \rightarrow 0$  for  $h \neq 0$ . A similar statement holds for  $c^{-q}R\left((1 + c\gamma)^{-1}ch_2\right)$  as well. Since  $1 + c\gamma \rightarrow 1$ , we obtain  $D(h) = D(h_2)$  which proves the first claim. The second claim is then an immediate consequence since  $D(0) = 0$  by homogeneity.

2. Suppose the claim were false. We could then find a sequence  $h_m \rightarrow 0$  with  $T(e + h_m) > T(e)$ . Rewrite  $h_m$  as  $\gamma_m e + h_{m2}$  with  $h_{m2} = \Pi_{\text{span}(e)^\perp} h_m$ . Clearly,  $\gamma_m \rightarrow 0$  would have to hold, implying  $1 + \gamma_m > 0$  for all sufficiently large  $m$ . Using  $G_X$ -invariance we obtain  $T(e + h_m) = T(e + (1 + \gamma_m)^{-1} h_{m2})$  for all large  $m$ . In particular, we conclude that  $h_{m2} \neq 0$  would have to hold for all large  $m$ . Applying (12) to the r.h.s. of the preceding equation we thus obtain for all large  $m$

$$0 < T(e + h_m) - T(e) = D\left((1 + \gamma_m)^{-1} h_{m2}\right) + R\left((1 + \gamma_m)^{-1} h_{m2}\right).$$

Using homogeneity of  $D$  we then have for all large  $m$

$$0 < D(h_{m2}/\|h_{m2}\|) + R\left((1 + \gamma_m)^{-1} h_{m2}\right) / \left\| (1 + \gamma_m)^{-1} h_{m2} \right\|^q = D(h_{m2}/\|h_{m2}\|) + o(1).$$

Note that  $h_{m2}/\|h_{m2}\|$  is an element of the compact set  $S^{n-1} \cap \text{span}(e)^\perp$  on which  $D$  is continuous and negative. Hence, the r.h.s. of the preceding display is eventually bounded from above by zero, a contradiction. ■

Inspection of the proof of Part 1 of the preceding lemma shows that this proof in fact does not make use of the property that  $D$  does not vanish on all of  $\text{span}(e)^\perp$ .

**Proof of Corollary 2.21:** 1. Clearly  $T_B(e) > \kappa \geq \lambda_1(B)$  implies  $e \notin \text{span}(X)$  in view of the definition of  $T_B$ . In view of the assumption on  $\kappa$ , the rejection region satisfies  $\emptyset \neq \Phi_{B,\kappa} \neq \mathbb{R}^n$ . Consequently  $T_B(e) > \kappa$  implies  $e \notin \text{bd}(\Phi_{B,\kappa})$ , cf. Proposition 2.11. But  $e \in \Phi_{B,\kappa}$  clearly holds, implying that  $e \in \text{int}(\Phi_{B,\kappa})$ . The result then follows immediately from Theorem 2.7 combined with the observation that  $\mathbf{1}_{\Phi_{B,\kappa}}$  is continuous at  $e$  if and only if  $e \notin \text{bd}(\Phi_{B,\kappa})$ .

2. Since  $e \notin \text{span}(X)$  by assumption, we conclude similarly as above that  $T_B(e) < \kappa$  implies  $e \notin \text{bd}(\Phi_{B,\kappa})$ . But  $e \notin \Phi_{B,\kappa}$  clearly holds, implying that  $e \notin \text{cl}(\Phi_{B,\kappa})$ . As before, the result then follows from Theorem 2.7. ■

**Proof of Corollary 2.22:** Observe that (11) is satisfied for  $\mathbf{1}_{\Phi_{B,\kappa}}$  since  $T_B$  is  $G_X$ -invariant and  $e \in \text{span}(X)$  by assumption. Hence, all assumptions of Part B of Theorem 2.16 are satisfied and thus the existence and the form of the limit follows. If  $\kappa > \lambda_1(B)$  the test  $\mathbf{1}_{\Phi_{B,\kappa}}$  is neither  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to zero nor  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to one, whereas  $\mathbf{1}_{\Phi_{B,\kappa}}$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere equal to one if  $\kappa = \lambda_1(B)$  as discussed in Remark 2.12. Part B of Theorem 2.16 and Remark 2.17(iv) then deliver the remaining claims. ■

**Proof of Corollary 2.23:** All assumptions for Part 2 of Theorem 2.18 (including the elliptic symmetry assumption) except for (12) are obviously satisfied. We first consider the situation of Part 1 of the corollary: That  $\lambda = T_B(e)$  follows immediately from  $e \notin \text{span}(X)$  and the definition of  $T_B$ . Furthermore, it was shown in Example 2.4 that (12) holds with  $q = 2$  and  $D$  given by (20), and that  $D$  satisfies all conditions required in Theorem 2.18. Applying the second part of Theorem 2.18 with  $q = 2$  then immediately gives (21). Furthermore, observe that

$$\Lambda' (C'_X B C_X - \lambda C'_X C_X) \Lambda = A' (C'_X B C_X - \lambda C'_X C_X) A \quad (48)$$

where  $A = \Lambda + ee'$  is nonsingular (cf. the proof of Theorem 2.16). By the general assumptions we have  $\lambda < \lambda_{n-k}(B)$ . If now  $\lambda > \lambda_1(B)$  holds, we see that the matrix in (48) is not equal to the zero matrix and is indefinite. Consequently, the r.h.s. of (21) is strictly between zero and one. In case  $\lambda = \lambda_1(B)$  the matrix in (48) is again not equal to the zero matrix, but is now nonnegative definite, which shows that the r.h.s. of (21) equals 1.

Next consider the situation of Part 2 of the corollary: As shown in Example 2.4, now condition (12) holds with  $q = 1$  and  $D$  given by (19), and  $D$  satisfies all conditions required in Theorem 2.18.

Applying the second part of Theorem 2.18 now with  $q = 1$  then immediately gives (22). The claim regarding (22) falling into  $(0, 1)$  then follows immediately from Remark 2.20(iii), while the final claim follows from this in conjunction with Remark 2.20(ii). The claim in parenthesis follows from the second part of Theorem 2.18 and the following observation: Note that  $\Lambda U'_0 e = 0$  implies that

$$a_1 := \left( e' C'_X B C_X - \|C_X e\|^{-2} (e' C'_X B C_X e) e' C'_X C_X \right) \Lambda$$

and  $e' U_0$  are orthogonal. Furthermore,  $a_1 \neq 0$  since the matrix in parentheses in the definition of  $a_1$  does not vanish on all of  $\text{span}(e)^\perp$  (see Example 2.4). Since also  $e' U_0 \neq 0$ , we conclude that  $a_1$  and  $e' U_0$  are not collinear.

Finally, Part 3 of the corollary follows immediately from Part 3 of Theorem 2.18 observing that  $q = 1$  as shown by Example 2.4. ■

**Proof of Lemma 2.25:** Let  $\kappa$  be a real number such that  $\kappa < T(e)$  and  $0 < |\kappa - T(e)| < \delta$ . Then  $e \in \Phi_\kappa$  and  $e \notin \text{bd}(\Phi_\kappa)$  hold, implying that  $e \in \text{int}(\Phi_\kappa)$ . Theorem 2.7 and Remark 2.8 then entail  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_\kappa) = 1$ . If  $\kappa < T(e)$  but  $|\kappa - T(e)| \geq \delta$  the same conclusion can be drawn since  $\Phi_{\kappa_1} \supseteq \Phi_{\kappa_2}$  for  $\kappa_1 \leq \kappa_2$ . Therefore, we have  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_\kappa) = 1$  for every  $\kappa < T(e)$ . Next, let  $\kappa$  be a real number such that  $\kappa > T(e)$  and  $|\kappa - T(e)| < \delta$  hold. This implies  $e \notin \Phi_\kappa$  and  $e \notin \text{bd}(\Phi_\kappa)$ , and hence  $e \notin \text{cl}(\Phi_\kappa)$ . Theorem 2.7 and Remark 2.8 now give  $\lim_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_\kappa) = 0$  for those values of  $\kappa$ . Monotonicity of  $\Phi_\kappa$  w.r.t.  $\kappa$  shows that this relation must hold for all  $\kappa > T(e)$ . From (23) and the just established results we obtain

$$\alpha^*(T) = \begin{cases} \inf_{\kappa < T(e)} P_{0,1,0}(\Phi_\kappa) & \text{if } \liminf_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{T(e)}) = 0, \\ \inf_{\kappa \leq T(e)} P_{0,1,0}(\Phi_\kappa) & \text{if } \liminf_{\rho \rightarrow a} P_{0,1,\rho}(\Phi_{T(e)}) > 0. \end{cases}$$

The function  $\kappa \mapsto P_{0,1,0}(\Phi_\kappa)$  is precisely one minus the cumulative distribution function of  $P_{0,1,0} \circ T$ , and hence is continuous at  $T(e)$  by assumption. Since it is clearly also decreasing in  $\kappa$ , we may conclude that

$$\alpha^*(T) = \inf_{\kappa < T(e)} P_{0,1,0}(\Phi_\kappa) = \inf_{\kappa \leq T(e)} P_{0,1,0}(\Phi_\kappa) = P_{0,1,0}(\Phi_{T(e)}).$$

Finally note that the claim in parenthesis is an immediate consequence of the second part of Proposition 2.11. ■

**Lemma C.4.** *Suppose that  $Q$  is a probability measure on  $\mathbb{R}^n$  which is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$ . Let  $T_B$  be given by (8).*

1. *Then the support of  $Q \circ T_B$  is contained in  $[\lambda_1(B), \lambda_{n-k}(B)]$ . Furthermore, if  $\lambda_1(B) < \lambda_{n-k}(B)$ , the cumulative distribution function of  $Q \circ T_B$  is continuous on the real line.*
2. *If the density of  $Q$  is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set, then the support of  $Q \circ T_B$  is  $[\lambda_1(B), \lambda_{n-k}(B)]$ .*

**Proof:** 1. Observe that the image of  $\mathbb{R}^{n-k} \setminus \{0\}$  under the map  $v \mapsto v' B v / v' v$  is  $[\lambda_1(B), \lambda_{n-k}(B)]$ . Because  $T_B$  is defined to be  $\lambda_1(B)$  on  $\text{span}(X)$ , it follows that the range of  $T_B$  is contained in  $[\lambda_1(B), \lambda_{n-k}(B)]$ , implying that the support of  $Q \circ T_B$  is contained in the same interval. [We note for later use that the range of  $T_B$  actually coincides with all of  $[\lambda_1(B), \lambda_{n-k}(B)]$ , because  $C_X : \mathbb{R}^n \rightarrow \mathbb{R}^{n-k}$  is surjective.] Next assume that  $\lambda_1(B) < \lambda_{n-k}(B)$ . To prove the continuity of the cumulative distribution function  $c \mapsto (Q \circ T_B)((-\infty, c])$  it suffices to show that  $(Q \circ T_B)(\{c\})$

is equal to zero for every  $c \in \mathbb{R}$ . Note that  $Q(\text{span}(X)) = 0$  since  $Q$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$  and  $k < n$  holds. Consequently, we have for every  $c \in \mathbb{R}$

$$(Q \circ T_B)(\{c\}) = Q(\{y \in \mathbb{R}^n : y' C'_X(B - cI_{n-k})C_X y = 0\}).$$

To show that  $(Q \circ T_B)(\{c\}) = 0$  it suffices to show that  $\mu_{\mathbb{R}^n}(\{y \in \mathbb{R}^n : y' C'_X(B - cI_{n-k})C_X y = 0\}) = 0$ . The set under consideration is obviously an algebraic set. Hence, it is a  $\mu_{\mathbb{R}^n}$ -null set if we can show that the quadratic form in the definition of this set does not vanish everywhere. Suppose the contrary, i.e.,  $y' C'_X(B - cI_{n-k})C_X y = 0$  for every  $y \in \mathbb{R}^n$  would hold. Because  $C_X : \mathbb{R}^n \rightarrow \mathbb{R}^{n-k}$  is surjective,  $v'(B - cI_{n-k})v = 0$  for every  $v \in \mathbb{R}^{n-k}$  would have to hold. Since  $B - cI_{n-k}$  is symmetric, this would imply  $B - cI_{n-k} = 0$ , contradicting  $\lambda_1(B) < \lambda_{n-k}(B)$ . This establishes  $(Q \circ T_B)(\{c\}) = 0$ .

2. If  $\lambda_1(B) = \lambda_{n-k}(B)$  this is trivial. Hence assume  $\lambda_1(B) < \lambda_{n-k}(B)$ . Let  $\lambda$  be an element in the interior of  $[\lambda_1(B), \lambda_{n-k}(B)]$  and let  $\varepsilon > 0$  arbitrary. Without loss of generality assume that  $\varepsilon$  is sufficiently small such that  $(\lambda - \varepsilon, \lambda + \varepsilon) \subseteq [\lambda_1(B), \lambda_{n-k}(B)]$ . Let  $y \in \mathbb{R}^n$  be such that  $T_B(y) = \lambda$ . Such an  $y$  exists, because the range of  $T_B$  is all of  $[\lambda_1(B), \lambda_{n-k}(B)]$  as noted in the proof of Part 1. But then  $y \notin \text{span}(X)$  must hold (since  $\lambda > \lambda_1(B)$ ), and hence  $T_B$  is continuous at  $y$ . Consequently, there is an open ball that is mapped into  $(\lambda - \varepsilon, \lambda + \varepsilon)$  by  $T_B$ . By  $G_X$ -invariance of  $T_B$  any open neighborhood of the origin contains such a ball. Because  $Q$  has a density that is almost everywhere positive on a sufficiently small open neighborhood of the origin, we see that  $Q \circ T_B$  puts positive mass on  $(\lambda - \varepsilon, \lambda + \varepsilon)$ . ■

**Proof of Proposition 2.26:** 1. Noting that  $e \notin \text{span}(X)$  and that  $T_B$  is continuous on  $\mathbb{R}^n \setminus \text{span}(X)$ , we may use Lemma 2.25 in conjunction with the preceding Lemma C.4 with  $Q = P_{0,1,0}$  to conclude that  $\alpha^*(T_B) = P_{0,1,0}(\Phi_{B,T_B(e)})$ . Note that this quantity can also be written as  $1 - (P_{0,1,0} \circ T_B)((-\infty, T_B(e)])$ . Thus  $\alpha^*(T_B) = 0$  is equivalent to the cumulative distribution function of  $T_B$  under  $P_{0,1,0}$  being equal to one when evaluated at  $T_B(e)$ . Lemma C.4 implies that this is in turn equivalent to  $T_B(e) = \lambda_{n-k}(B)$  (since  $T_B(e) > \lambda_{n-k}(B)$  is clearly impossible). But  $T_B(e) = \lambda_{n-k}(B)$  is clearly equivalent to  $C_X e \in \text{Eig}(B, \lambda_{n-k}(B))$ . This proves the first claim of Part 1. Next observe that for every  $\kappa \in (-\infty, \lambda_{n-k}(B))$  the assumptions on  $P_{0,1,0}$  together with Part 2 of Lemma C.4 imply  $P_{0,1,0}(\Phi_{B,\kappa}) > 0 = \alpha^*(T_B)$ . The second claim then follows from Lemma 2.25. For the claim in parenthesis see Remark 2.12.

2. By the same reasoning as in the proof of Part 1 we see that  $\alpha^*(T_B) = 1$  is then equivalent to  $T_B(e) = \lambda_1(B)$ . Since  $e \notin \text{span}(X)$  by assumption, this is in turn equivalent to  $C_X e \in \text{Eig}(B, \lambda_1(B))$ . This proves the first claim of Part 2. The second claim follows directly from Lemma 2.25 because  $P_{0,1,0}(\Phi_{B,\kappa}) < 1 = \alpha^*(T_B)$  holds for  $\kappa$  in the specified range in view of Lemma C.4 and the assumptions on  $P_{0,1,0}$ . The remaining claims follow from Remark 2.12.

3. The first claim is obvious in light of Parts 1 and 2, and the remaining claims follows from Lemma C.4 and Lemma 2.25. ■

**Proof of Proposition 2.28:** The test is obviously invariant w.r.t.  $G_X$ , and the additional invariance condition (11) in Theorem 2.16 is satisfied because of  $e \in \text{span}(X)$ . If  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$ , the rejection region as well as its complement have positive  $\mu_{\mathbb{R}^n}$ -measure, whereas  $\Phi_{B,\kappa}$  is  $\mathbb{R}^n$  or the complement of a  $\mu_{\mathbb{R}^n}$ -null set in case  $\kappa \leq \lambda_1(B)$ , see Remark 2.12. The second claim then follows from Theorem 2.16, Part A.3, and Remark 2.17(i) in case  $\kappa \in (\lambda_1(B), \lambda_{n-k}(B))$ , and is obvious otherwise. Now, the just established claim implies that  $\alpha^*(T_B)$  is not larger than  $\inf\{P_{0,1,0}(\Phi_{B,\kappa}) : \kappa < \lambda_{n-k}(B)\} = P_{0,1,0}(\{T_B \geq \lambda_{n-k}(B)\})$ . Since the set  $\{T_B \geq \lambda_{n-k}(B)\}$  is a  $\mu_{\mathbb{R}^n}$ -null set (as  $\lambda_1(B) < \lambda_{n-k}(B)$  is assumed) and since  $P_{0,1,0}$  is absolutely continuous by the assumptions of the lemma,  $\alpha^*(T_B) = 0$  then follows. The claim in parentheses is trivial. ■

**Lemma C.5.** Let  $A \in \mathbb{R}^{n \times n}$  be a symmetric positive definite matrix and let  $\delta \geq 0$ . Then, the following statements are equivalent:

- (i)  $C_X A C'_X = \delta I_{n-k}$  for some matrix  $C_X$  satisfying  $C_X C'_X = I_{n-k}$  and  $C'_X C_X = \Pi_{\text{span}(X)^\perp}$ ,
- (ii)  $C_X A C'_X = \delta I_{n-k}$  for any matrix  $C_X$  satisfying  $C_X C'_X = I_{n-k}$  and  $C'_X C_X = \Pi_{\text{span}(X)^\perp}$ ,
- (iii)  $\Pi_{\text{span}(X)^\perp} A \Pi_{\text{span}(X)^\perp} = \delta \Pi_{\text{span}(X)^\perp}$ ,
- (iv) there exists a matrix  $D$  such that  $DD' = A$  and  $\Pi_{\text{span}(X)^\perp} D = \delta^{1/2} \Pi_{\text{span}(X)^\perp}$  holds.

**Proof:** That (i), (ii) and (iii) are equivalent is obvious from the relations  $C_X C'_X = I_{n-k}$  and  $C'_X C_X = \Pi_{\text{span}(X)^\perp}$ . That (iv) implies (iii) is obvious. To see that (iii) implies (iv), note that  $\Pi_{\text{span}(X)^\perp}$  is symmetric and idempotent, thus

$$\Pi_{\text{span}(X)^\perp} A^{1/2} A^{1/2} \Pi_{\text{span}(X)^\perp} = (\delta^{1/2} \Pi_{\text{span}(X)^\perp})(\delta^{1/2} \Pi_{\text{span}(X)^\perp})'.$$

In other words,  $\Pi_{\text{span}(X)^\perp} A^{1/2}$  and  $\delta^{1/2} \Pi_{\text{span}(X)^\perp}$  are both square roots of the same matrix, which implies existence of an orthogonal matrix,  $U$  say, such that

$$\Pi_{\text{span}(X)^\perp} A^{1/2} U = \delta^{1/2} \Pi_{\text{span}(X)^\perp}.$$

Setting  $D = A^{1/2} U$  then completes the proof. ■

**Proof of Theorem 2.30:** 1. Clearly, as  $\Sigma(\rho^*)$  is positive definite, we must have  $C_X \Sigma(\rho^*) C'_X = \delta I_{n-k}$  with  $\delta = \delta(\rho^*) > 0$ . By Lemma C.5, there exists an  $n \times n$  matrix  $D = D(\rho^*)$  such that  $DD' = \Sigma(\rho^*)$  and  $\Pi_{\text{span}(X)^\perp} D = \delta^{1/2} \Pi_{\text{span}(X)^\perp}$ . Since  $D$  is a square root of  $\Sigma(\rho^*)$  there exists an orthogonal matrix  $U(\rho^*)$  such that  $K(\rho^*) = DU(\rho^*)$ . Now observe that

$$\Pi_{\text{span}(X)^\perp} (X\beta + \sigma K(\rho^*)z) = \sigma \Pi_{\text{span}(X)^\perp} K(\rho^*)z = \sigma \delta^{1/2} \Pi_{\text{span}(X)^\perp} U(\rho^*)z. \quad (49)$$

This immediately gives the last equality in (24). Now, if  $\Pi_{\text{span}(X)^\perp} (X\beta + \sigma K(\rho^*)z) \neq 0$ , then we can use the equation in the previous display to obtain

$$\mathcal{I}_X(X\beta + \sigma K(\rho^*)z) = \langle \Pi_{\text{span}(X)^\perp} U(\rho^*)z / \|\Pi_{\text{span}(X)^\perp} U(\rho^*)z\| \rangle = \mathcal{I}_X(U(\rho^*)z)$$

and

$$\mathcal{I}_X^+(X\beta + \sigma K(\rho^*)z) = \Pi_{\text{span}(X)^\perp} U(\rho^*)z / \|\Pi_{\text{span}(X)^\perp} U(\rho^*)z\| = \mathcal{I}_X^+(U(\rho^*)z).$$

If  $\Pi_{\text{span}(X)^\perp} (X\beta + \sigma K(\rho^*)z) = 0$ , then also  $\Pi_{\text{span}(X)^\perp} U(\rho^*)z = 0$  in view of (49). Hence, also in this case we obtain  $\mathcal{I}_X(X\beta + \sigma K(\rho^*)z) = \mathcal{I}_X(U(\rho^*)z)$  and  $\mathcal{I}_X^+(X\beta + \sigma K(\rho^*)z) = \mathcal{I}_X^+(U(\rho^*)z)$ . This proves Part 1.

2. Observe that under the assumption on  $\mathfrak{P}$  the distribution of  $\mathcal{I}_X^1$  under  $P_{\beta, \sigma, \rho^*}$  is the distribution of  $\mathcal{I}_X^1(X\beta + \sigma L(\rho^*)\mathbf{z}) = \mathcal{I}_X^1(\sigma \delta^{1/2} U(\rho^*)\mathbf{z})$  (upon choosing  $K(\rho^*) = L(\rho^*)$ ) which coincides with the distribution of  $\mathcal{I}_X^1(\sigma \delta^{1/2} \mathbf{z})$  by the implied spherical symmetry of the distribution of  $\mathbf{z}$ . But clearly, the distribution of  $\mathcal{I}_X^1(\sigma \delta^{1/2} \mathbf{z})$  coincides with the distribution of  $\mathcal{I}_X^1(\sigma \delta^{1/2} L(0)\mathbf{z})$  by spherical symmetry and since  $\Sigma(0) = I_n$  implies that  $L(0)$  is an orthogonal matrix. In turn, the distribution of  $\mathcal{I}_X^1(\sigma \delta^{1/2} L(0)\mathbf{z})$  coincides with the distribution of  $\mathcal{I}_X^1$  under  $P_{0, \sigma \delta^{1/2}, 0}$  since  $\mathfrak{P}$ , in particular, satisfies Assumption 3. This proves that  $P_{\beta, \sigma, \rho^*} \circ \mathcal{I}_X^1 = P_{0, \sigma \delta^{1/2}(\rho^*), 0} \circ \mathcal{I}_X^1$ . That  $P_{\beta, \sigma, 0} \circ \mathcal{I}_X^1 = P_{0, \sigma, 0} \circ \mathcal{I}_X^1$  can be proved in the same way observing that  $C_X \Sigma(0) C'_X = C_X C'_X = I_{n-k}$ . [Alternatively, it follows immediately from  $G_X^1$ -invariance and the fact that the distribution of  $\mathbf{z}$  does not depend on  $\beta$ .] The proofs for the corresponding statements regarding the distributions of  $\mathcal{I}_X$  and  $\mathcal{I}_X^+$  are analogous. Since every other invariant statistic can be represented as a function of  $\mathcal{I}_X$ ,  $\mathcal{I}_X^+$ , and  $\mathcal{I}_X^1$ , respectively, the second claim of Part 2 follows. The third claim is now obvious. ■

## D Proofs for Sections 4.1 and 4.2

**Proof of Lemma 4.2:** Suppose  $\sigma_1^2 \Sigma_{SEM}(\rho_1) = \sigma_2^2 \Sigma_{SEM}(\rho_2)$  and set  $\tau = \sigma_2^2 / \sigma_1^2$ . This implies

$$(\tau - 1)I_n = (\tau\rho_1 - \rho_2)(W' + W) + (\rho_2^2 - \tau\rho_1^2)W'W. \quad (50)$$

If  $\tau = 1$  inspection of the diagonal elements in (50) shows that all diagonal elements of  $(\rho_2^2 - \rho_1^2)W'W$  must be zero, which is only possible if  $\rho_2^2 - \rho_1^2 = 0$  since  $W$  can not be the zero matrix. But then we arrive at  $\rho_1 = \rho_2$  and  $\sigma_1 = \sigma_2$ . Now suppose  $\tau \neq 1$  would hold. Then inspection of the diagonal elements in (50) shows that the diagonal elements of  $W'W$  are all identical equal to  $b > 0$ , say, and must satisfy  $\tau - 1 = (\rho_2^2 - \tau\rho_1^2)b$ , which can equivalently be written as

$$\tau(1 + \rho_1^2 b) = 1 + \rho_2^2 b$$

Furthermore, multiplying (50) by  $f'_{\max}$  from the left and by  $f_{\max}$  from the right and noting that  $f'_{\max}f_{\max} = 1$  holds, gives after a rearrangement

$$\tau(1 - \rho_1\lambda_{\max})^2 = (1 - \rho_2\lambda_{\max})^2.$$

Expressing  $\tau$  from the last equation (note that  $1 - \rho_1\lambda_{\max} > 0$ ), and substituting into the last but one equation gives

$$(1 + \rho_1^2 b) / (1 - \rho_1\lambda_{\max})^2 = (1 + \rho_2^2 b) / (1 - \rho_2\lambda_{\max})^2.$$

But the function  $\rho \mapsto (1 + \rho^2 b) / (1 - \rho\lambda_{\max})^2$  is obviously strictly increasing on  $[0, \lambda_{\max}^{-1})$  since  $b > 0$  holds. This gives  $\rho_1 = \rho_2$  and consequently also  $\tau = 1$  would hold, a contradiction. ■

**Proof of Lemma 4.3:** Clearly,  $\Sigma_{SEM}^{-1}((\lambda_{\max}^{-1}) -) = (I_n - \lambda_{\max}^{-1}W')(I_n - \lambda_{\max}^{-1}W)$  and its kernel equals the kernel of  $I_n - \lambda_{\max}^{-1}W$  which obviously contains  $f_{\max}$  and which is one-dimensional by the assumptions on  $W$ . Therefore the kernel equals  $\text{span}(f_{\max})$ , which together with Lemma 2.5 proves the first claim. To prove the second claim we need to show that  $\Lambda$  in the formulation of the lemma is well-defined, is injective when restricted to  $\text{span}(f_{\max})^\perp$ , and satisfies

$$\Pi_{\text{span}(f_{\max})^\perp} (I_n - \rho W)^{-1} \rightarrow \Lambda \quad (51)$$

for  $\rho \rightarrow \lambda_{\max}^{-1}$  with  $\rho \in [0, \lambda_{\max}^{-1})$ . Observe that for every  $0 \leq \rho < \lambda_{\max}^{-1}$  we can find a  $\delta(\rho) < 1$  such that  $\rho\lambda_{\max} < \delta(\rho)$  holds. Noting that  $\rho\lambda_{\max}$  is the spectral radius of  $\rho W$  by our assumptions on  $W$ , we can conclude that  $\|(\rho W)^j\|^{1/j} \rightarrow \rho\lambda_{\max} < \delta(\rho)$  for  $j \rightarrow \infty$  (where  $\|\cdot\|$  denotes an arbitrary matrix norm), cf. Horn and Johnson (1985), Corollary 5.6.14. But then it follows that  $(I_n - \rho W)^{-1}$  can be written as the norm-convergent series  $\sum_{j=0}^{\infty} \rho^j W^j$  for every  $0 \leq \rho < \lambda_{\max}^{-1}$ . Thus we obtain

$$\Pi_{\text{span}(f_{\max})^\perp} (I_n - \rho W)^{-1} = \Pi_{\text{span}(f_{\max})^\perp} \sum_{j=0}^{\infty} \rho^j W^j = \Pi_{\text{span}(f_{\max})^\perp} + \sum_{j=1}^{\infty} \rho^j \Pi_{\text{span}(f_{\max})^\perp} W^j. \quad (52)$$

Let  $g_2, \dots, g_n$  be an orthonormal basis of  $\text{span}(f_{\max})^\perp$  and define the  $n \times (n-1)$  matrix  $U_2 = (g_2, \dots, g_n)$ . Then  $U = (f_{\max} : U_2)$  is an orthogonal matrix. Set  $D = U' W U$  and observe that  $D$  takes the form

$$D = \begin{pmatrix} \lambda_{\max} & b' \\ 0 & A \end{pmatrix}.$$



For later use we note that  $\lambda_{\max}$  is not an eigenvalue of  $A$  since the eigenvalues of  $D$  and  $W$  coincide, since the eigenvalues of  $D$  are made up of  $\lambda_{\max}$  and the eigenvalues of  $A$ , and because  $\lambda_{\max}$  has algebraic multiplicity 1 by assumption. Now clearly

$$\Pi_{\text{span}(f_{\max})^\perp} W^j = \Pi_{\text{span}(f_{\max})^\perp} U D^j U' = U_2 A^j U_2' \quad (53)$$

holds for  $j \geq 1$ , which implies

$$\Pi_{\text{span}(f_{\max})^\perp} W^j = U_2 A^j U_2' = (U_2 A U_2')^j = (\Pi_{\text{span}(f_{\max})^\perp} W)^j$$

for  $j \geq 1$ . Consequently,

$$\begin{aligned} \Pi_{\text{span}(f_{\max})^\perp} (I_n - \rho W)^{-1} &= \Pi_{\text{span}(f_{\max})^\perp} + \sum_{j=1}^{\infty} \rho^j (\Pi_{\text{span}(f_{\max})^\perp} W)^j \\ &= -\Pi_{\text{span}(f_{\max})} + \sum_{j=0}^{\infty} \rho^j (\Pi_{\text{span}(f_{\max})^\perp} W)^j \\ &= (I_n - \rho \Pi_{\text{span}(f_{\max})^\perp} W)^{-1} - \Pi_{\text{span}(f_{\max})}, \end{aligned} \quad (54)$$

observing that the infinite sum in the second line of (54) is norm-convergent because of (52), and thus necessarily equals the inverse matrix in the last line of (54). Because  $\lambda_{\max}$  is not an eigenvalue of  $\Pi_{\text{span}(f_{\max})^\perp} W$  in view of (53) with  $j = 1$ , the matrix  $I_n - \lambda_{\max}^{-1} \Pi_{\text{span}(f_{\max})^\perp} W$  is invertible, showing that  $\Lambda$  is well-defined. Furthermore, from (54) we see that (51) indeed holds. Finally,  $\Lambda$  is injective on  $\text{span}(f_{\max})^\perp$  since  $\Lambda$  coincides with  $(I_n - \lambda_{\max}^{-1} \Pi_{\text{span}(f_{\max})^\perp} W)^{-1}$  on this subspace. ■

**Proof of Lemma 4.4:** The first claim is an obvious consequence of the maintained assumptions for the SEM. The second claim follows from Proposition 2.6 together with the already established first claim, since Assumption 1 holds for the SEM as shown in Lemma 4.3. ■

**Proof of Corollary 4.5:** Parts 1-3 follow from combining Lemmata 4.3, 4.4, Theorem 2.7, Remark 2.8, Theorem 2.16, and Remark 2.17(i), noting that here  $L = L_*$ . Part 4 is then a simple consequence of Part 3 in view of Proposition 2.11, Remark 2.12, and Remark 2.17(iv); cf. also the proof of Corollary 2.21. ■

**Proof of Corollary 4.7:** The first part follows immediately from Part 1 of Corollary 2.23. The second part follows from Part 3 of the same corollary if we can verify that the additional condition assumed there is satisfied. First observe that  $f_{\max}$  is an eigenvector of  $I_n - \rho W$  to the eigenvalue  $1 - \rho \lambda_{\max}$  and that  $I_n - \rho W$  is nonsingular for  $0 \leq \rho < \lambda_{\max}^{-1}$ . Because  $I_n - \rho W$  is symmetric,  $f_{\max}$  is then also an eigenvector of  $\Sigma_{SEM}(\rho) = (I_n - \rho W)^{-2}$  with eigenvalue  $(1 - \rho \lambda_{\max})^{-2}$ . Next observe that  $\Pi_{\text{span}(f_{\max})} = f_{\max} f_{\max}'$ . But then we have for  $0 \leq \rho < \lambda_{\max}^{-1}$

$$\begin{aligned} \Pi_{\text{span}(f_{\max})^\perp} \Sigma_{SEM}(\rho) \Pi_{\text{span}(f_{\max})} &= \Pi_{\text{span}(f_{\max})^\perp} (I_n - \rho W)^{-2} f_{\max} f_{\max}' \\ &= (1 - \rho \lambda_{\max})^{-2} \Pi_{\text{span}(f_{\max})^\perp} f_{\max} f_{\max}' = 0. \end{aligned}$$

■

**Proof of Proposition 4.9:** Lemmata 4.3 and 4.4 show that Assumptions 1 and 2 are satisfied. In view of the assumptions of the proposition,  $P_{0,1,0}$  is clearly absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$  with a density that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set, and  $e = f_{\max} \notin \text{span}(X)$  is trivially satisfied since  $k = 0$ . Obviously,  $W + W'$  is not a

multiple of the identity matrix. [If it were, inspection of the diagonal elements shows that  $W + W'$  would have to be the zero matrix. However, this is also impossible since  $f'_{\max}(W + W')f_{\max} = 2f'_{\max}Wf_{\max} = 2\lambda_{\max} > 0$ .] Also,  $-\Sigma_{SEM}^{-1}(\bar{\rho})$  cannot be a multiple of the identity matrix in view of Lemma 4.2. Hence, in both cases we have  $\lambda_1(B) < \lambda_n(B)$ . Proposition 2.26 and the observation that the rejection probabilities are monotonically decreasing in  $\kappa$  now establishes the first claim of the proposition. It remains to show that  $f_{\max} \in \text{Eig}(B, \lambda_n(B))$  is equivalent to  $f_{\max}$  being an eigenvector of  $W'$  under the additional assumptions on  $W$ . We may assume that  $f_{\max}$  is entrywise positive. We argue here similarly as in the proof of Proposition 1 in Martellosio (2010). Consider first the case where  $B = -\Sigma_{SEM}^{-1}(\bar{\rho})$ . If  $f_{\max} \in \text{Eig}(B, \lambda_n(B))$  then

$$\lambda_n(B)f_{\max} = Bf_{\max} = -(1 - \bar{\rho}\lambda_{\max})(I - \bar{\rho}W')f_{\max}$$

from which it follows that  $f_{\max}$  is an eigenvector of  $W'$ . Conversely, if  $f_{\max}$  is an eigenvector of  $W'$  then  $f_{\max}$  is easily seen to be also an eigenvector of  $B = -\Sigma_{SEM}^{-1}(\bar{\rho})$  and hence also of  $-B^{-1}$ . Now,  $-B^{-1} = \Sigma_{SEM}(\bar{\rho})$  is an entrywise positive matrix by a result in Gantmacher (1959), p. 69. Consequently, the eigenspace corresponding to its largest eigenvalue is one-dimensional and is spanned by a unique normalized and entrywise positive eigenvector  $g$ , say. Since  $-B^{-1}$  is symmetric and  $f_{\max}$  is an entrywise positive eigenvector of  $-B^{-1}$ , it must correspond to the largest eigenvalue of  $-B^{-1}$  (because otherwise it would have to be orthogonal to  $g$ , which is impossible as  $f_{\max}$  and  $g$  are both entrywise nonnegative). Hence,  $f_{\max} \in \text{Eig}(-B^{-1}, \lambda_n(-B^{-1})) = \text{Eig}(B, \lambda_n(B))$ . Next consider the case  $B = W + W'$ . As before,  $f_{\max} \in \text{Eig}(B, \lambda_n(B))$  implies that  $f_{\max}$  is an eigenvector of  $W'$ . Conversely,  $f_{\max}$  being an eigenvector of  $W'$  implies that  $f_{\max}$  is an eigenvector of  $B$ . Since  $W + W'$  is symmetric, entrywise nonnegative, and irreducible (since  $W$  is so) the same argument as in the first case can be applied. ■

**Proof of Proposition 4.10:** As in the proof of Proposition 4.9 it follows that Assumptions 1 and 2 are satisfied and that  $P_{0,1,0}$  is absolutely continuous w.r.t.  $\mu_{\mathbb{R}^n}$  with a density that is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set. By assumption  $f_{\max} \notin \text{span}(X)$  holds. Consider first case (ii): Observe that the eigenspaces of  $C_X \Sigma_{SEM}(\rho) C'_X$  and  $C_X \lambda_n^{-1}(\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho) C'_X$  are identical. By Assumption 1 we have for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ ,

$$C_X \lambda_n^{-1}(\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho) C'_X \rightarrow C_X f_{\max} f'_{\max} C'_X$$

the limiting matrix being a matrix of rank exactly equal to 1 since  $C_X f_{\max} \neq 0$  by the assumption  $f_{\max} \notin \text{span}(X)$ . Hence, its largest eigenvalue is positive and has algebraic multiplicity 1, while all other eigenvalues are zero. It follows from Tyler (1981), p. 726, Lemma 2.1, that then the eigenspace corresponding to the largest eigenvalue of  $C_X \lambda_n^{-1}(\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho) C'_X$  (and thus the eigenspace corresponding to the largest eigenvalue of  $C_X \Sigma_{SEM}(\rho) C'_X$ ) converges to the eigenspace of the limiting matrix corresponding to its largest eigenvalue (in the sense that the corresponding projection matrices onto these spaces converge). The latter space is obviously given by  $\text{span}(C_X f_{\max})$ . Because the eigenspaces of  $C_X \Sigma_{SEM}(\rho) C'_X$  corresponding to the largest eigenvalue are independent of  $\rho$  by assumption, it follows that these eigenspaces all coincide with  $\text{span}(C_X f_{\max})$ . Consequently, also  $\text{Eig}(B, \lambda_{n-k}(B)) = \text{span}(C_X f_{\max})$  holds for  $B = -(C_X \Sigma_{SEM}(\bar{\rho}) C'_X)^{-1}$ . In particular,  $\lambda_1(B) < \lambda_{n-k}(B)$  follows, as  $n - k > 1$  has been assumed. The result now follows from the first part of Proposition 2.26.

Next consider case (i): By assumption  $\lambda_1(B) < \lambda_{n-k}(B)$  holds. Hence we may apply the first part of Proposition 2.26 and it remains to show that  $C_X f_{\max}$  belongs to  $\text{Eig}(B, \lambda_{n-k}(B))$ . Now, observe that  $D(\rho) := C_X (\Sigma_{SEM}(\rho) - I_n) C'_X / \rho \rightarrow B$  for  $\rho \rightarrow 0$ ,  $\rho > 0$ . Because  $C_X f_{\max}$  is

an eigenvector of  $C_X \Sigma_{SEM}(\rho) C'_X$  corresponding to its largest eigenvalue,  $\nu(\rho)$  say, as was shown above, it is also an eigenvector of  $D(\rho)$  corresponding to its largest eigenvalue, namely  $(\nu(\rho) - 1)/\rho$ . Because  $D(\rho) \rightarrow B$  for  $\rho \rightarrow 0$ , it follows that  $C_X f_{\max}$  is an eigenvector of  $B$  corresponding to the limit of  $(\nu(\rho) - 1)/\rho$ , which necessarily then needs to coincide with the largest eigenvalue of  $B$ . ■

**Proof of Theorem 4.12:** Observe that the covariance matrix of  $\mathbf{y}$  under  $P_{\beta, \sigma, \rho}$  is given by  $\sigma^2 \Sigma_{SEM}(\rho)$ . Now, for  $\rho \in [0, \lambda_{\max}^{-1})$  we have

$$\lambda_{\max}^{-1/2} (\Sigma_{SEM}(\rho)) (I_n - \rho W)^{-1} = [\lambda_{\max}^{-1} (\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho)]^{1/2} U(\rho),$$

for a suitable orthogonal  $n \times n$  matrix  $U(\rho)$ . From Lemma 4.3 we know that  $\lambda_{\max}^{-1} (\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho)$  converges to  $f_{\max} f'_{\max}$  as  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$ . Continuity and uniqueness of the symmetric square root hence gives

$$[\lambda_{\max}^{-1} (\Sigma_{SEM}(\rho)) \Sigma_{SEM}(\rho)]^{1/2} \rightarrow (f_{\max} f'_{\max})^{1/2} = f_{\max} f'_{\max}.$$

Now, let  $\rho_m \rightarrow \lambda_{\max}^{-1}$ ,  $\rho_m \in [0, \lambda_{\max}^{-1})$  be an arbitrary sequence. Then we can always find a subsequence  $m'$  such that along this subsequence  $U(\rho_m)$  converges to an orthogonal matrix  $U$ . Consequently,  $\lambda_{\max}^{-1/2} (\Sigma_{SEM}(\rho_{m'})) (I_n - \rho_{m'} W)^{-1}$  converges to  $f_{\max} f'_{\max} U$ . Under  $P_{\beta, \sigma, \rho_{m'}}$  the random vector  $\mathbf{y} / \|\mathbf{y}\|$  clearly has the same distribution as

$$\lambda_{\max}^{-1/2} (\Sigma_{SEM}(\rho_{m'})) (I_n - \rho_{m'} W)^{-1} (X\beta + \sigma \mathbf{z}) / \left\| \lambda_{\max}^{-1/2} (\Sigma_{SEM}(\rho_{m'})) (I_n - \rho_{m'} W)^{-1} (X\beta + \sigma \mathbf{z}) \right\|$$

where  $\mathbf{z}$  is a fixed random vector distributed according to the distribution of  $\boldsymbol{\varepsilon}$ , which is independent of the parameters by assumption. Observing that the random variable  $f'_{\max} U (X\beta + \sigma \mathbf{z})$  is almost surely nonzero by the assumption on the distribution of  $\boldsymbol{\varepsilon}$ , the expression in the preceding display is now seen to converge in distribution as  $m' \rightarrow \infty$  to

$$f_{\max} f'_{\max} U (X\beta + \sigma \mathbf{z}) / \|f_{\max} f'_{\max} U (X\beta + \sigma \mathbf{z})\| = \mathbf{c} f_{\max}$$

where  $\mathbf{c}$  is a random variable with values in  $\{-1, 1\}$ . It then follows from the continuous mapping theorem that  $\mathcal{I}_{0, \zeta_{f_{\max}}}(\mathbf{y})$  converges in distribution under  $P_{\beta, \sigma, \rho_{m'}}$  to  $\zeta_{f_{\max}}(\mathbf{c} f_{\max}) = f_{\max}$ . In other words,  $P_{\beta, \sigma, \rho_{m'}} \circ \mathcal{I}_{0, \zeta_{f_{\max}}}$  converges weakly to pointmass  $\delta_{f_{\max}}$ . Now observe that

$$\begin{aligned} E_{\beta, \sigma, \rho_{m'}}(\varphi) &= \int_{\mathbb{R}^n} \varphi(y) dP_{\beta, \sigma, \rho_{m'}}(y) = \int_{\mathbb{R}^n} \varphi(\mathcal{I}_{0, \zeta_{f_{\max}}}(y)) dP_{\beta, \sigma, \rho_{m'}}(y) \\ &= \int_{\mathbb{R}^n} \varphi(y) d(P_{\beta, \sigma, \rho_{m'}} \circ \mathcal{I}_{0, \zeta_{f_{\max}}})(y). \end{aligned}$$

But the r.h.s. of the preceding display converges to  $\varphi(f_{\max})$  because  $P_{\beta, \sigma, \rho_{m'}} \circ \mathcal{I}_{0, \zeta_{f_{\max}}}$  converges weakly to pointmass  $\delta_{f_{\max}}$  and because  $\varphi$  is bounded and is continuous at  $f_{\max}$ , cf. Theorem 30.12 in Bauer (2001). A standard subsequence argument then shows that the limit of  $E_{\beta, \sigma, \rho}(\varphi)$  for  $\rho \rightarrow \lambda_{\max}^{-1}$ ,  $\rho \in [0, \lambda_{\max}^{-1})$  is as claimed. The second claim is an immediate consequence of the first one. ■

**Proof of Lemma 4.13:** From (30) we obtain  $W' \Pi_{\text{span}(X)^\perp} = \lambda \Pi_{\text{span}(X)^\perp}$ . For  $0 \leq \rho < \lambda_{\max}^{-1}$  we thus obtain

$$(I_n - \rho W')^{-1} \Pi_{\text{span}(X)^\perp} = (1 - \rho \lambda)^{-1} \Pi_{\text{span}(X)^\perp},$$

which after transposition establishes the first claim. An immediate consequence of the first claim is

$$\Pi_{\text{span}(X)^\perp} \Sigma_{SEM}(\rho) \Pi_{\text{span}(X)^\perp} = (1 - \rho \lambda)^{-2} \Pi_{\text{span}(X)^\perp}$$

which establishes the second claim in view of Lemma C.5. ■

## E Auxiliary Results

**Lemma E.1.** *Let  $\mathbf{z}$  be a random  $n$ -vector with a density,  $p$  say, w.r.t.  $\mu_{\mathbb{R}^n}$ . Then  $\mathbf{s} = \mathbf{z}/\|\mathbf{z}\|$  is well-defined with probability 1 and has a density,  $\bar{p}$  say, w.r.t. the uniform probability measure  $v_{S^{n-1}}$  on  $S^{n-1}$ . The density  $\bar{p}$  satisfies*

$$\bar{p}(s) = c \int_{(0,\infty)} p(rs) r^{n-1} d\mu_{(0,\infty)}(r)$$

*$v_{S^{n-1}}$ -almost everywhere, where  $c = 2\pi^{n/2}/\Gamma(n/2)$ . Furthermore, if  $p$  is positive on an open neighborhood of the origin except possibly for a  $\mu_{\mathbb{R}^n}$ -null set (which is, in particular, the case if  $p$  is positive  $\mu_{\mathbb{R}^n}$ -almost everywhere), then  $\bar{p}$  is positive  $v_{S^{n-1}}$ -almost everywhere.*

**Proof:** Let  $B$  be a Borel set in  $S^{n-1}$  and let  $\chi : \mathbb{R}^n \setminus \{0\} \rightarrow S^{n-1}$  be given by  $\chi(z) = z/\|z\|$ . Then

$$\Pr(\mathbf{s} \in B) = \Pr(\mathbf{z} \in \chi^{-1}(B)) = \int_{\mathbb{R}^n \setminus \{0\}} \mathbf{1}_{\chi^{-1}(B)}(z) p(z) dz = \int_{(0,\infty) \times S^{n-1}} \mathbf{1}_{\chi^{-1}(B)}(rs) p(rs) dH(r, s)$$

where  $H$  is the pushforward measure of  $\mu_{\mathbb{R}^n}$  (restricted to  $\mathbb{R}^n \setminus \{0\}$ ) under the map  $z \mapsto (\|z\|, z/\|z\|)$ . But  $H$  is nothing else than the product of the measure on  $(0, \infty)$  with Lebesgue density  $r^{n-1}$  and the surface measure  $cv_{S^{n-1}}$  on  $S^{n-1}$  where  $c$  is given in the lemma (cf. Stroock (1999)). In view of Tonelli's theorem (observe all functions involved are nonnegative) and since  $\mathbf{1}_{\chi^{-1}(B)}(rs) = \mathbf{1}_{\chi^{-1}(B)}(s) = \mathbf{1}_B(s)$  clearly holds for  $s \in S^{n-1}$ , we obtain

$$\Pr(\mathbf{s} \in B) = \int_{S^{n-1}} \mathbf{1}_B(s) \left( c \int_{(0,\infty)} p(rs) r^{n-1} d\mu_{(0,\infty)}(r) \right) dv_{S^{n-1}}(s),$$

which establishes the claims except for the last one. We next prove the final claim. First, observe that for every Borel set  $B$  in  $S^{n-1}$  we have  $v_{S^{n-1}}(B) > 0$  if and only if  $\mu_{\mathbb{R}^n}(\chi^{-1}(B)) > 0$ . [This is seen as follows: Specializing what has been proved so far to the case where  $\mathbf{z}$  follows a standard Gaussian distribution, shows that in this case  $\mathbf{s}$  is uniformly distributed on  $S^{n-1}$ . Hence,  $v_{S^{n-1}}(B) = \Pr(\mathbf{s} \in B) = \Pr(\mathbf{z} \in \chi^{-1}(B))$ . But then the equivalence of the Gaussian measure with  $\mu_{\mathbb{R}^n}$  establishes that  $v_{S^{n-1}}(B) > 0$  if and only if  $\mu_{\mathbb{R}^n}(\chi^{-1}(B)) > 0$ .] Let now  $B$  satisfy  $v_{S^{n-1}}(B) > 0$ . Clearly,  $\Pr(\mathbf{s} \in B) = \Pr(\mathbf{z} \in \chi^{-1}(B)) \geq \Pr(\mathbf{z} \in \chi^{-1}(B) \cap V)$  where  $V$  is an open neighborhood of the origin on which  $p$  is positive  $\mu_{\mathbb{R}^n}$ -almost everywhere. But then we must have  $\mu_{\mathbb{R}^n}(\chi^{-1}(B) \cap V) > 0$ , because  $\mu_{\mathbb{R}^n}(\chi^{-1}(B)) > 0$  follows as a consequence of  $v_{S^{n-1}}(B) > 0$  as just shown above and because  $\chi^{-1}(B)$  can be written as a countable union of the sets  $j(\chi^{-1}(B) \cap V)$  with  $j \in \mathbb{N}$ . By the assumption on  $p$  we can now conclude that  $\Pr(\mathbf{z} \in \chi^{-1}(B) \cap V) > 0$  holds. Hence, we have established that  $\Pr(\mathbf{s} \in B) > 0$  holds whenever  $v_{S^{n-1}}(B) > 0$  is satisfied. ■

**Remark E.2.** (i) In the proof we have shown that for  $B$ , a Borel subset of the unit sphere, we have  $v_{S^{n-1}}(B) > 0$  if and only if  $\mu_{\mathbb{R}^n}(\chi^{-1}(B)) > 0$ , a fact that we shall freely use in various places.

(ii) Let  $\mathbf{z}$  be a random  $n$ -vector such that  $\Pr(\mathbf{z} = 0) = 0$ . Assume that  $\mathbf{z}/\|\mathbf{z}\|$  has a density w.r.t.  $v_{S^{n-1}}$  (which is, in particular, the case if  $\mathbf{z}$  is spherically symmetric). Let  $A$  be a  $G_0^+$ -invariant Borel set in  $\mathbb{R}^n$  with  $\mu_{\mathbb{R}^n}(A) = 0$ . Then  $\Pr(\mathbf{z} \in A) = 0$  holds. To see this use  $G_0^+$ -invariance and the fact that  $\mathbf{z}$  has no atom at the origin to obtain  $\Pr(\mathbf{z} \in A) = \Pr(\mathbf{z} \in A \setminus \{0\}) = \Pr(\mathbf{z}/\|\mathbf{z}\| \in A \setminus \{0\}) = \Pr(\mathbf{z}/\|\mathbf{z}\| \in B)$ , where  $B = \chi(A \setminus \{0\})$ . Note that  $B$  is a Borel subset

of  $S^{n-1}$  satisfying  $\chi^{-1}(B) = A \setminus \{0\}$ . Hence  $\mu_{\mathbb{R}^n}(\chi^{-1}(B)) = 0$  holds. But then  $v_{S^{n-1}}(B) = 0$  by what was shown in (i). Since  $\mathbf{s} = \mathbf{z}/\|\mathbf{z}\|$  possesses a density w.r.t.  $v_{S^{n-1}}$  by assumption, we conclude that  $\Pr(\mathbf{z}/\|\mathbf{z}\| \in B) = 0$ , and thus also  $\Pr(\mathbf{z} \in A) = 0$  must hold.

(iii) Let  $\mathbf{z}$  be as in (ii) and let  $A$  be a  $G_X^+$ -invariant Borel set in  $\mathbb{R}^n$  with  $\mu_{\mathbb{R}^n}(A) = 0$ . Then for every  $\beta \in \mathbb{R}^n$ ,  $0 < \sigma < \infty$ , and every nonsingular  $n \times n$  matrix  $L$  we have  $\Pr(X\beta + \sigma L\mathbf{z} \in A) = \Pr(L\mathbf{z} \in A) = \Pr(\mathbf{z} \in L^{-1}(A)) = 0$  in view of (ii) since  $L^{-1}(A)$  is a  $G_0^+$ -invariant  $\mu_{\mathbb{R}^n}$ -null set.

**Lemma E.3.** *Let  $\mathbf{z}$  be a random  $n$ -vector satisfying  $\Pr(\mathbf{z} = 0) = 0$ . Then  $\mathbf{s} = \mathbf{z}/\|\mathbf{z}\|$  is well-defined with probability 1. Assume further that the distribution of  $\mathbf{s}$  has a density,  $g$  say, w.r.t.  $v_{S^{n-1}}$ . Suppose  $\mathbf{r}$  is a random variable taking values in  $(0, \infty)$  that is independent of  $\mathbf{z}/\|\mathbf{z}\|$  and that has a density,  $h$  say, w.r.t.  $\mu_{(0, \infty)}$ . Define  $\mathbf{z}^\dagger = \mathbf{r}\mathbf{z}/\|\mathbf{z}\|$  on the event  $\mathbf{z} \neq \mathbf{0}$  and assign arbitrary values to  $\mathbf{z}^\dagger$  on the event  $\mathbf{z} = \mathbf{0}$  in a measurable way. Then, the following holds:*

1.  $\Pr(\mathbf{z}^\dagger = 0) = 0$  and  $\mathbf{z}^\dagger/\|\mathbf{z}^\dagger\| = \mathbf{z}/\|\mathbf{z}\|$  for  $\mathbf{z} \neq 0$ ,  $\mathbf{z}^\dagger \neq 0$ .
2.  $\mathbf{z}^\dagger$  possesses a density  $g^\dagger$  w.r.t. Lebesgue measure  $\mu_{\mathbb{R}^n}$  which is given by

$$g^\dagger(z) = \begin{cases} c^{-1}g(z/\|z\|) \frac{h(\|z\|)}{\|z\|^{n-1}} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0, \end{cases}$$

where  $c$  has been given in Lemma E.1.

3. If  $g$  is  $v_{S^{n-1}}$ -almost everywhere continuous and  $h$  is  $\mu_{(0, \infty)}$ -almost everywhere continuous, then  $g^\dagger$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere continuous.
4. If  $g$  is  $v_{S^{n-1}}$ -almost everywhere positive and  $h$  is  $\mu_{(0, \infty)}$ -almost everywhere positive, then  $g^\dagger$  is  $\mu_{\mathbb{R}^n}$ -almost everywhere positive.
5. If  $g$  is constant  $v_{S^{n-1}}$ -almost everywhere [which is, in particular, the case if  $\mathbf{z}$  is spherically symmetric] and if  $\mathbf{r}$  is distributed as the square root of a  $\chi^2$ -distributed random variable with  $n$  degrees of freedom, then  $\mathbf{z}^\dagger$  is Gaussian with mean zero and covariance matrix  $I_n$ .

**Proof:** Part 1 is obvious. To prove Part 2 we denote the distribution of  $\mathbf{z}/\|\mathbf{z}\|$  by  $G$  and the distribution of  $\mathbf{r}$  by  $H$ . Because  $\mathbf{z}/\|\mathbf{z}\|$  and  $\mathbf{r}$  are independent, the joint distribution of  $\mathbf{z}/\|\mathbf{z}\|$  and  $\mathbf{r}$  on  $S^{n-1} \times (0, \infty)$ , equipped with the product  $\sigma$ -field, is given by the product measure  $G \otimes H$ . Therefore, the distribution of  $\mathbf{z}^\dagger$  is the push-forward measure of  $G \otimes H$  under the mapping  $m(s, r) = rs$ . Hence for every  $A \in \mathcal{B}(\mathbb{R}^n)$  we have, using Tonelli's theorem and the fact that  $G$  and  $H$  have densities  $g$  and  $h$ , respectively, that

$$\begin{aligned} \Pr(\mathbf{z}^\dagger \in A) &= \int_{S^{n-1} \times (0, \infty)} \mathbf{1}_A(rs) d(G \otimes H)(s, r) = \int_{(0, \infty)} \int_{S^{n-1}} \mathbf{1}_A(rs) dG(s) dH(r) \\ &= \int_{(0, \infty)} \int_{S^{n-1}} \mathbf{1}_A(rs) g(s) dv_{S^{n-1}}(s) h(r) d\mu_{(0, \infty)}(r) \\ &= \int_{(0, \infty)} r^{n-1} \int_{S^{n-1}} \mathbf{1}_A(rs) g(s) r^{1-n} h(r) dv_{S^{n-1}}(s) d\mu_{(0, \infty)}(r) \\ &= \int_{(0, \infty)} r^{n-1} \int_{S^{n-1}} f(rs) dv_{S^{n-1}}(s) d\mu_{(0, \infty)}(r), \end{aligned}$$

where for  $x \in \mathbb{R}^n$  the function  $f$  is given by

$$f(x) = \begin{cases} \mathbf{1}_A(x)g(x/\|x\|)\|x\|^{1-n}h(\|x\|) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Since  $f$  is clearly a non-negative and Borel-measurable function, we can apply Theorem 5.2.2 in Stroock (1999) to see that

$$\begin{aligned} \Pr(\mathbf{z}^\dagger \in A) &= \int_{(0,\infty)} r^{n-1} \int_{S^{n-1}} f(rs) dv_{S^{n-1}}(s) d\mu_{(0,\infty)}(r) \\ &= \int_{\mathbb{R}^n} c^{-1} f(x) d\mu_{\mathbb{R}^n}(x) = \int_{\mathbb{R}^n} \mathbf{1}_A(x) g^\dagger(x) d\mu_{\mathbb{R}^n}(x). \end{aligned}$$

This establishes the second part of the lemma. To prove the third part denote by  $D_{g^\dagger} \subseteq \mathbb{R}^n$ ,  $D_g \subseteq S^{n-1}$  and  $D_h \subseteq (0, \infty)$  the discontinuity points of  $g^\dagger$ ,  $g$ , and  $h$ , respectively, which are measurable. Using Part 2 of the lemma we see that  $x \neq 0$ ,  $x/\|x\| \in \mathbb{R}^n \setminus D_g$ , and  $\|x\| \in \mathbb{R}^n \setminus D_h$  imply  $x \in \mathbb{R}^n \setminus D_{g^\dagger}$ . Therefore, negating the statement, we see that  $\mathbf{1}_{D_{g^\dagger}}(x) \leq \mathbf{1}_{\{0\}}(x) + \mathbf{1}_{D_g}(x/\|x\|) + \mathbf{1}_{D_h}(\|x\|)$  must hold which implies

$$\mu_{\mathbb{R}^n}(D_{g^\dagger}) = \int_{\mathbb{R}^n} \mathbf{1}_{D_{g^\dagger}}(x) d\mu_{\mathbb{R}^n}(x) \leq \int_{\mathbb{R}^n} \mathbf{1}_{D_g}(x/\|x\|) d\mu_{\mathbb{R}^n}(x) + \int_{\mathbb{R}^n} \mathbf{1}_{D_h}(\|x\|) d\mu_{\mathbb{R}^n}(x). \quad (55)$$

Using again Theorem 5.2.2 in Stroock (1999) we see that

$$\begin{aligned} \int_{\mathbb{R}^n} \mathbf{1}_{D_g}(x/\|x\|) d\mu_{\mathbb{R}^n}(x) &= \int_{(0,\infty)} r^{n-1} \int_{S^{n-1}} \mathbf{1}_{D_g}(s) c dv_{S^{n-1}}(s) d\mu_{(0,\infty)}(r) \\ &= \int_{(0,\infty)} c v_{S^{n-1}}(D_g) r^{n-1} d\mu_{(0,\infty)}(r) = 0, \end{aligned}$$

because  $v_{S^{n-1}}(D_g) = 0$  holds by assumption. Similarly, we obtain

$$\int_{\mathbb{R}^n} \mathbf{1}_{D_h}(\|x\|) d\mu_{\mathbb{R}^n}(x) = \int_{S^{n-1}} \int_{(0,\infty)} r^{n-1} \mathbf{1}_{D_h}(r) d\mu_{(0,\infty)}(r) c dv_{S^{n-1}}(s) = 0,$$

because the inner integral is zero as a consequence of the assumption that  $\mu_{(0,\infty)}(D_h) = 0$ . Together with Equation (55) the last two displays establish  $\mu_{\mathbb{R}^n}(D_{g^\dagger}) = 0$ . To prove Part 4 denote by  $Z_{g^\dagger} \subseteq \mathbb{R}^n$ ,  $Z_g \subseteq S^{n-1}$ , and  $Z_h \subseteq (0, \infty)$  the zero sets of  $g^\dagger$ ,  $g$ , and  $h$ , respectively, which are obviously measurable. Replacing  $D_{g^\dagger}$ ,  $D_g$ , and  $D_h$  with  $Z_{g^\dagger}$ ,  $Z_g$ , and  $Z_h$ , respectively, in the argument used above then establishes Part 4. To prove the last part, we observe that  $g$  being constant  $v_{S^{n-1}}$ -almost everywhere implies that  $\mathbf{z}/\|\mathbf{z}\|$  is uniformly distributed on  $S^{n-1}$ . Since  $\mathbf{z}/\|\mathbf{z}\|$  is independent of  $\mathbf{r}$ , which is distributed as the square root of a  $\chi^2$  with  $n$  degrees of freedom, it is now obvious that  $\mathbf{z}^\dagger$  is Gaussian with mean zero and covariance matrix  $I_n$ . ■

**Remark E.4.** As long as we are only concerned with distributional properties of  $\mathbf{z}$  we can assume w.l.o.g. that the probability space supporting  $\mathbf{z}$  is rich enough to allow independent random variables  $\mathbf{r}$  that have the required properties. In particular, we can then always choose  $\mathbf{r}$  such that the density is simultaneously  $\mu_{(0,\infty)}$ -almost everywhere continuous and  $\mu_{(0,\infty)}$ -almost everywhere positive (e.g., by choosing  $\mathbf{r}$  to follow a  $\chi^2$ -distribution).

## References

- ANSELIN, L. (2001). Spatial econometrics. In *A companion to theoretical econometrics*. Blackwell Companions Contemp. Econ., Blackwell, Malden, MA, 310–330.
- ARNOLD, S. F. (1979). Linear models with exchangeably distributed errors. *Journal of the American Statistical Association*, **74** 194–199.
- BARTELS, R. (1992). On the power function of the Durbin-Watson test. *Journal of Econometrics*, **51** 101–112.
- BAUER, H. (2001). *Measure and integration theory*, vol. 26 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin.
- CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11** 368–385.
- GANTMACHER, F. R. (1959). *The theory of matrices, vol 2*. AMS Chelsea Publishing, Providence, RI.
- HORN, R. A. and JOHNSON, C. R. (1985). *Matrix analysis*. Cambridge University Press, Cambridge.
- KADIYALA, K. R. (1970). Testing for the independence of regression disturbances. *Econometrica*, **38** 97–117.
- KARIYA, T. (1980). Note on a condition for equality of sample variances in a linear model. *Journal of the American Statistical Association*, **75** 701–703.
- KING, M. L. (1985). A point optimal test for autoregressive disturbances. *Journal of Econometrics*, **27** 21–37.
- KING, M. L. (1987). Testing for autocorrelation in linear regression models: a survey. In *Specification analysis in the linear model*. Internat. Lib. Econom., Routledge & Kegan Paul, London, 19–73.
- KING, M. L. and HILLIER, G. H. (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **47** 98–102.
- KLEIBER, C. and KRÄMER, W. (2005). Finite-sample power of the Durbin-Watson test against fractionally integrated disturbances. *Econometrics Journal*, **8** 406–417.
- KRÄMER, W. (1985). The power of the Durbin-Watson test for regressions without an intercept. *Journal of Econometrics*, **28** 363–370.
- KRÄMER, W. (2005). Finite sample power of Cliff-Ord-type tests for spatial disturbance correlation in linear regression. *Journal of Statistical Planning and Inference*, **128** 489–496.
- KRÄMER, W. and ZEISEL, H. (1990). Finite sample power of linear regression autocorrelation tests. *Journal of Econometrics*, **43** 363–372.
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing statistical hypotheses*. 3rd ed. Springer Texts in Statistics, Springer, New York.
- LÖBUS, J.-U. and RITTER, L. (2000). The limiting power of the Durbin-Watson test. *Communications in Statistics-Theory and Methods*, **29** 2665–2676.
- MARTELLOSIO, F. (2010). Power properties of invariant tests for spatial autocorrelation in linear regression. *Econometric Theory*, **26** 152–186.

- MARTELLOSIO, F. (2011a). Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Statistics & Probability Letters*, **81** 1285 – 1291.
- MARTELLOSIO, F. (2011b). Nontestability of equal weights spatial dependence. *Econometric Theory*, **27** 1369–1375.
- MARTELLOSIO, F. (2012). Testing for spatial autocorrelation: The regressors that make the power disappear. *Econometric Reviews*, **31** 215–240.
- MYNBAEV, K. (2012). Distributions escaping to infinity and the limiting power of the Cliff-Ord test for autocorrelation. *ISRN Probability and Statistics*, **2012**.
- PREINERSTORFER, D. (2014). How to avoid the zero-power trap in testing for correlation. *Working paper, in preparation*.
- PREINERSTORFER, D. and PÖTSCHER, B. M. (2013). On size and power of heteroskedasticity and autocorrelation robust tests. *Econometric Theory*, *forthcoming*.
- SMALL, J. P. (1993). The limiting power of point optimal autocorrelation tests. *Communications in Statistics-Theory and Methods*, **22** 3907–3916.
- STROOCK, D. W. (1999). *A concise introduction to the theory of integration*. 3rd ed. Birkhäuser Boston Inc., Boston, MA.
- TILLMAN, J. (1975). The power of the Durbin-Watson test. *Econometrica*, **43** 959–974.
- TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Annals of Statistics*, **9** pp. 725–736.
- ZEISEL, H. (1989). On the power of the Durbin-Watson test under high autocorrelation. *Communications in Statistics-Theory and Methods*, **18** 3907–3916.